

INTERNATIONAL SERIES IN PURE AND APPLIED
PHYSICS

LEE A. DUBRIDGE, CONSULTING EDITOR

Advisory Editorial Committee:

E. U. CONDON, GEORGE R. HARRISON, ELMER HUTCHISSON,
K. K. DARROW

Wave Propagation in
Periodic Structures

*The quality of the materials used in
the manufacture of this book is gov-
erned by continued postwar shortages.*

INTERNATIONAL SERIES IN PURE AND APPLIED PHYSICS

LEE A. DuBRIDGE, *Consulting Editor*

- BACHER AND GOUDSMIT—ATOMIC ENERGY STATES
BITTER—INTRODUCTION TO FERROMAGNETISM
BRILLOUIN—WAVE PROPAGATION IN PERIODIC STRUCTURES
CADY—PIEZOELECTRICITY
CLARK—APPLIED X-RAYS
CONDON AND MORSE—QUANTUM MECHANICS
CURTIS—ELECTRICAL MEASUREMENTS
DAVEY—CRYSTAL STRUCTURE AND ITS APPLICATIONS
EDWARDS—ANALYTIC AND VECTOR MECHANICS
ELDRIDGE—THE PHYSICAL BASIS OF THINGS
HARDY AND PERRIN—THE PRINCIPLES OF OPTICS
HARNWELL—PRINCIPLES OF ELECTRICITY AND ELECTRO-
MAGNETISM
HARNWELL AND LIVINGOOD—EXPERIMENTAL ATOMIC PHYSICS
HOUSTON—PRINCIPLES OF MATHEMATICAL PHYSICS
HUGHES AND DuBRIDGE—PHOTOELECTRIC PHENOMENA
HUND—HIGH-FREQUENCY MEASUREMENTS
PHENOMENA IN HIGH-FREQUENCY SYSTEMS
KEMBLE—THE FUNDAMENTAL PRINCIPLES OF QUANTUM
MECHANICS
KENNARD—KINETIC THEORY OF GASES
KOLLER—THE PHYSICS OF ELECTRON TUBES
MORSE—VIBRATION AND SOUND
MUSKAT—THE FLOW OF HOMOGENEOUS FLUIDS THROUGH
POROUS MEDIA
PAULING AND GOUDSMIT—THE STRUCTURE OF LINE SPECTRA
RICHTMYER—INTRODUCTION TO MODERN PHYSICS
RUARK AND UREY—ATOMS, MOLECULES AND QUANTA
SLATER—INTRODUCTION TO CHEMICAL PHYSICS
SLATER AND FRANK—INTRODUCTION TO THEORETICAL
PHYSICS
SMYTHE—STATIC AND DYNAMIC ELECTRICITY
WHITE—INTRODUCTION TO ATOMIC SPECTRA
WILLIAMS—MAGNETIC PHENOMENA

Wave Propagation in Periodic Structures

Electric Filters and Crystal Lattices

BY

LÉON BRILLOUIN

Professor, Collège de France, Paris

Applied Mathematics Group; Columbia University

Vice President, Ecole Libre des Hautes Etudes, New York

First Edition

Second Impression

New York

London

M c G R A W - H I L L B O O K C O M P A N Y , I N C .

1946

WAVE PROPAGATION IN PERIODIC STRUCTURES

COPYRIGHT, 1946, BY THE
MCGRAW-HILL BOOK COMPANY, INC.

PRINTED IN THE UNITED STATES OF AMERICA

*All rights reserved. This book, or
parts thereof, may not be reproduced
in any form without permission of
the publishers.*

Dedicated to My Wife

WHOSE FAITH AND LOVE HAVE NEVER FAILED
THROUGH HARDSHIP AND SORROW

PREFACE

Some readers may be surprised or even disturbed at the mixture of problems assembled in this book. These problems actually extend from electrical engineering to electromagnetism and wave mechanics of the spinning electron, but the link connecting this variety of problems will soon be discovered in their common mathematical background.

Waves always behave in a similar way, whether they are longitudinal or transverse, elastic or electric. Scientists of the last century always kept this idea in mind. When Lord Kelvin built up his model for a dispersive medium or when Lord Rayleigh discovered radiation pressure, they never failed to try the same methods again and again on all conceivable types of waves. This general philosophy of wave propagation, forgotten for a time, has been strongly revived in the last decade and represents the backbone of this book.

All problems discussed deal with periodic structures of various kinds, and they all lead to similar results: these structures, be they electric lines or crystal lattices, behave like band-pass filters. If energy dissipation is omitted, there is a sharp distinction between frequency bands exhibiting wave propagation without attenuation (passing bands) and those showing attenuation and no propagation (stopping bands). These general properties are defined for an infinite unbounded medium, but they bear a very close relation to *selective reflections* shown by a bounded medium. A wave striking from outside may be partly reflected from the surface, if the second medium is able to transmit the corresponding frequency. The amount of reflection depends upon how well the media are matched at their common boundary. But when the frequency falls inside a stopping band of the reflecting medium, there is no longer any matching problem; the wave cannot be transmitted, and hence it must be totally reflected. This same explanation applies to electric filters, rest rays, anomalous optical reflections, and selective reflection of X rays or electrons from a crystal. In the case of rest rays, the theory was developed

by M. Born and his school; for X rays, it corresponds to Bragg's reflections and P. P. Ewald's now classical investigations summarized in his book "*Kristalle und Roentgen Strahlen*" (Springer, 1923), and a paper in the *Annales de l'Institut Poincaré* (vol. 7, p. 79, 1938). Many practical examples of electric filters may be found in the treatises of K. S. Johnson and T. E. Shea, in the collection of books from the Bell Telephone Laboratories (van Nostrand). The general connection between stopping bands and selective reflection is exemplified in the definition of the *zones* for a crystal lattice, a theory first developed by the author in his original papers and in a book "*Quantenstatistik*" (Springer, 1931). A general discussion of the zone theory is found in the present book and will serve as an introduction to Mott and Jones, "*Theory of Metals and Alloys*" (Oxford, 1936), and to F. Seitz's "*The Modern Theory of Solids*" (McGraw-Hill, 1940), where the theory is applied to many practical discussions.

Apart from physical and engineering problems, the general theory developed in this book bears a close connection with many problems of applied mathematics, such as the Mathieu functions and Mathieu's and Hill's equations.

The author discussed these general problems in his lectures at the Collège de France (1937-1938) and at the University of Wisconsin (1942), when Mary Hewlett Payne very kindly proposed to write down her notes and to prepare them for publication. Circumstances resulted in great delays before this could be completed, and the author's present duties would never have permitted him to undertake such a work if Mrs. Payne had not made a really excellent record of his lectures, so that very few corrections and additions were necessary on her manuscript. Let her find here the author's very best thanks for her valuable collaboration.

LÉON BRILLOUIN.

NEW YORK, N. Y.,
January, 1946.

CONTENTS

	PAGE
PREFACE.	vii

Chapter I

ELASTIC WAVES IN A ONE-DIMENSIONAL LATTICE OF POINT MASSES: EARLY WORK AND INTRODUCTION

SECTION		
	1. Historical Background; Eighteenth Century.	1
	2. Historical Background; Nineteenth Century. Cauchy, Baden-Powell, and Kelvin	3
	3. Later Work on Models Similar to That Treated in Sec. 2	11

Chapter II

PROPAGATION OF WAVES ALONG ONE-DIMENSIONAL LATTICES. GENERAL RESULTS AND QUALITATIVE DISCUSSION

4. General Remarks	17
5. A Lattice of Free Particles	19
6. Longitudinal Vibration in a Row of Equidistant Coupled Oscillators	21
7. Longitudinal Vibrations in a Row of Diatomic Molecules	22

Chapter III

MATHEMATICAL TREATMENT OF A ONE-DIMENSIONAL LATTICE OF IDENTICAL PARTICLES

8. Equation of Motion of a One-dimensional Lattice of Identical Particles.	26
9. Rigorous Discussion for the Case of Interactions between Nearest Neighbors Only	31
10. Discussion of the Distance of Interaction.	33
11. The Low-pass Electric Filter	37
12. Analogies between Electrical and Mechanical Systems	40

Chapter IV

MATHEMATICAL TREATMENT OF
MORE COMPLICATED ONE-DIMENSIONAL LATTICES

SECTION	PAGE
13. Equations of Motion for the One-dimensional NaCl Lattice.	44
14. Electrical Analogue of the One-dimensional Diatomic Lattice.	47
15. Discussion of the One-dimensional NaCl Lattice. . . .	50
16. Transition from a Diatomic to a Monatomic Lattice. .	57
17. The One-dimensional Lattice of Polyatomic Molecules	65

Chapter V

ENERGY VELOCITY, ENERGY FLOW,
AND CHARACTERISTIC IMPEDANCE

18. General Discussion; Phase Velocity	69
19. A Theorem from the Theory of Complex Variables . . .	70
20. Energy Density, Energy Flow, and Energy Velocity . .	72
21. Group Velocity and Propagation of a Signal.	74
22. Preliminary Definition of Characteristic Impedance . .	80
23. Junction of Two Lattices.	85
24. General Definition of Characteristic Impedance	88

Chapter VI

TWO-DIMENSIONAL LATTICES

25. Direct and Reciprocal Lattices in Two Dimensions. . .	94
26. Doubly and Triply Periodic Functions	99
27. Zones in a Two-dimensional Lattice	102
28. Propagation of Waves in a Continuous Two-dimensional Medium with a Periodic Perturbation	107
29. The Exceptional Waves of Case 2 and Bragg Reflection	115
30. Transition Near the Discontinuity	118
31. Examples and Discussion of Zones in Two Dimensions.	121

Chapter VII

THREE-DIMENSIONAL LATTICE

32. Direct and Reciprocal Lattices in Three Dimensions . .	131
33. Zones in Three Dimensions and Bragg Reflection; Ewald's Construction	136

SECTION	PAGE
34. General Results for a Wave Propagating in a Three-dimensional Periodic Medium.	139
35. Waves in a Homogeneous Isotropic Medium with a Small Periodic Perturbation.	143
36. General Remarks on Waves in a Discontinuous Lattice	147
37. Some Examples of Zones in Three Dimensions.	148
38. Zones in the Direct Lattice; Principle of the Wigner-Seitz Method	155
39. Frequency Distribution for Waves in an Actual Crystal	157
40. The Energy of a Solid; the Characteristic Temperatures	164
41. Thermal Expansion and Entropy of a Solid Body.	167

Chapter VIII

MATHIEU'S EQUATION AND RELATED PROBLEMS

42. Mathieu's Equation	172
43. Mathieu Functions: General Discussion.	175
44. Hill's Equation with a Rectangular Curve	180
45. The Self-excited Oscillator	186
46. Free Electrons in Metals	190

Chapter IX

MATRICES AND THE PROPAGATION OF WAVES ALONG AN ELECTRIC LINE

47. General Remarks	193
48. Expressions for Energy.	197
49. Definition of a Four-terminal and Equation for Its Circuit.	200
50. Matrix Notation for a Four-terminal.	201
51. Combination of Two Four-terminals; Multiplication of Matrices	203
52. Inverse and Reversed Four-terminals and Transformations.	204
53. Four-terminal Matrices and the Group C_2	206
54. Surge, Iterative, or Characteristic Impedance of a Four-terminal	209
55. Propagation along a Line of Four-terminals.	211
56. Application of the Theory to a Reversible Four-terminal	214
57. Passing Bands and Attenuation in a Line of Four-terminals	216

SECTION	PAGE
58. Reflected Waves in a Line Terminated by an Impedance ζ_0	217
59. A Continuous Line Loaded with Two-terminals	218
60. A Continuous Line Loaded with Four-terminals	221

Chapter X

CONTINUOUS ELECTRIC LINES

61. Transition from a Line of Four-terminals to a Continuous Line	227
62. Examples of Four-terminal Representation of Continuous Lines	229
63. Application of Hill's Equation to a Continuous Line	232
64. Normalization of the Matrix (ϵ_{ij}) and the Pauli Matrices	234
65. Three-phase and Polyphase Lines	236
INDEX.	241

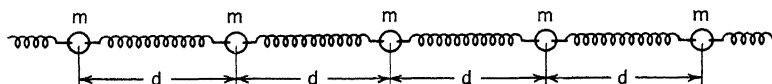
WAVE PROPAGATION IN PERIODIC STRUCTURES

CHAPTER I

ELASTIC WAVES IN A ONE-DIMENSIONAL LATTICE OF POINT MASSES: EARLY WORK AND INTRODUCTION

1. Historical Background; Eighteenth Century

The first work done on a one-dimensional lattice was that of Newton¹ in his attempt to derive a formula for the velocity of sound. Newton assumed that sound was propagated in air in the same manner in which an elastic wave would be propagated along a lattice of point masses. He assumed the simplest possible such lattice, *viz.*, one consisting of equal masses spaced



ELASTIC CONSTANT = e

FIG. 1.1.

equally along the direction of propagation (Fig. 1.1). Neighboring masses were assumed to attract one another with an elastic force with constant e . Taking m to be the mass of each of the particles and d to be the distance between neighboring particles in the state of equilibrium, Newton obtained for the velocity V of propagation of an elastic wave

$$V = d \sqrt{\frac{e}{m}} = \sqrt{\frac{ed}{\rho}} \quad \rho = \text{density} \quad (1.1)$$

To compare this result with the experimental value of the velocity of sound in air, Newton took ρ to be the density of air and ed to be the isothermal bulk modulus of air. The theoretical value thus computed was smaller than the experimental value. In 1822 Laplace pointed out that the expansions and condensations

¹ NEWTON, "Principia," Book II, 1686.

associated with sound waves take place adiabatically and that, therefore, the adiabatic elastic constant should be used instead of the isothermal. A computation using the adiabatic constant in Newton's formula gave excellent agreement with experiment. It should be mentioned that Newton's formula holds only for wave lengths large compared with d .

The reason why Newton considered the one-dimensional lattice of Fig. 1.1 was that at that time a continuous structure represented an insoluble problem, and nothing was known about partial differential equations. Hence, a model had to be chosen that would lead to a number of simultaneous equations of motion of the usual type.

The work on one-dimensional lattices was continued in a series of letters, starting in 1727, between John Bernoulli in Basel and his son Daniel in St. Petersburg at that time. They showed that a system with n point masses has n independent modes of vibration, *i.e.*, n proper frequencies. Later (1753), Daniel Bernoulli formulated the principle of superposition, which states that the general motion of a vibrating system is given by a superposition of its proper vibrations. This investigation may be said to form the beginning of theoretical physics as distinct from mechanics, in the sense that it is the first attempt to formulate laws for the motion of a system of particles rather than for that of a single particle. The principle of superposition is important, as it is a special case of a Fourier series, and in time it was extended to become a statement of Fourier's theorem.

The laws of vibrating strings were first discovered empirically, and in 1713 Taylor started a theoretical investigation. Euler's treatment of the continuous string by means of partial differential equations (1748) was much more complete. He took the string to be along the x axis and to be vibrating in some plane perpendicular to this axis. The result he obtained was that the displacement of the string was given by an arbitrary function of $(x \pm vt)$, where v is the velocity of propagation of the wave and t is the time, provided that the function satisfied certain continuity conditions. Euler's result started a controversy lasting until 1807. If one takes Euler's result and the principle of superposition together, one must conclude that any arbitrary function of $(x \pm vt)$ may be described by a superposition of sine and cosine functions, since it is well known that the proper vibrations of a

string are given by sine and cosine functions. This is, of course, merely a statement of Fourier's theorem, but Fourier's theorem was not proved until 1807, and to Euler's mind the theorem was almost an absurdity. Since he could not doubt the validity of his solution to the problem of the vibrating continuous string, Euler refused to accept the principle of superposition.

The Bernoullis had given the problem of the one-dimensional lattice of point masses a fairly complete treatment. Euler had solved the problem of the vibrating continuous string. The task of treating the continuous string as a limiting case of the one-dimensional lattice of point masses still remained. This problem was solved by Lagrange in 1759.

Lagrange followed Euler in refusing to accept the principle of superposition. This is very strange, since Lagrange's paper practically contains the principle of the Fourier series. A number of examples of trigonometric series were already known at the time, but it was not believed that such expansions could be used to represent any arbitrary function. In a paper on celestial mechanics, Clairaut (1754) actually had the proof, but it remained unnoticed; and it was left for Fourier to give the general statement and to emphasize its great practical and theoretical importance.

All this work at the end of the eighteenth century is most interesting since it cleared the way for a number of modern problems in theoretical physics as well as for pure mathematics:

Proper functions, proper values; first discovered in connection with proper vibrations of strings, plates, etc.

Fourier expansion; expansion in series of proper functions.

Partial differential equations.

Wave propagation.

Atomic theory of solids and crystal structure.

Lagrange's paper was often quoted by the famous electrical engineer Pupin, who discovered in Lagrange's theory the solution of an important problem of electrical engineering, the loaded cable.

2. Historical Background; Nineteenth Century. Cauchy, Baden-Powell, and Kelvin

In 1830, Cauchy used Newton's model in an attempt to account for dispersion of optical waves. Cauchy assumed that

light waves were just elastic waves of very high frequency. He obtained the result that for waves with wave length large compared with the distances between the point masses in the one-dimensional lattice, the velocity was independent of wave length. For shorter wave lengths and hence for higher frequencies, however, he showed that the velocity of propagation was a function of wave length. The result is correct for elastic waves; however, it did not agree quantitatively with values obtained experimentally for light waves.

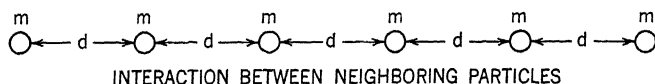
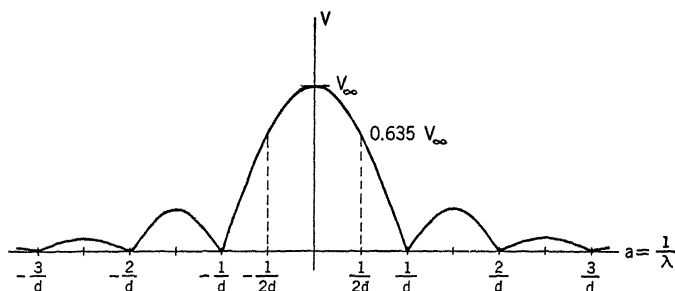


FIG. 2.1.

FIG. 2.2.—Wave velocity V as a function of a along the row of particles shown on Fig. 2.1.

In 1841, Baden-Powell computed the velocity of a wave propagating along one axis of a cubic lattice structure as a function of wave length. This is equivalent to considering a wave propagating along a one-dimensional lattice of point masses. Baden-Powell's lattice consisted of point masses of mass m spaced along a straight line at distance d from one another (see Fig. 2.1). Then he assumed each mass to be elastically bound to each of its neighbors with the restoring force the same for all masses. His equation for the propagation velocity V of the wave as a function of wave length is

$$V = V_{\infty} \frac{|\sin \pi d / \lambda|}{\pi d / \lambda} \quad (2.1)$$

where λ is the wave length and V_{∞} is the velocity for infinite wave length. The curve of V plotted against reciprocal wave length is shown in Fig. 2.2. It is evident that if velocity is a function

of wave length, the frequency must also be a function of the wave length. However, Baden-Powell neglected to consider the frequency as a function of the wave length and thus missed a very important point. The curve of velocity as a function of reciprocal wave length appears to be perfectly normal at the point $\lambda = 2d$; not so, however, for the frequency vs. reciprocal-wave-length curve. This point was noted by Kelvin, who gave a detailed discussion in 1881.¹

Kelvin assumed the same lattice that Baden-Powell treated (see Fig. 2.3). Let us number the particles in such a way that

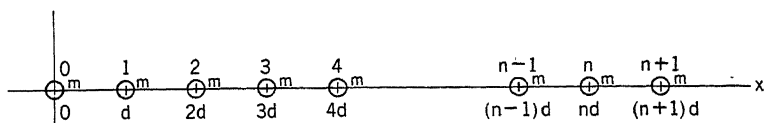


FIG. 2.3.

the x coordinate of the n th particle in its equilibrium position is given by

$$x_n = nd \quad (2.2)$$

In a sine wave, we obtain for y_n , the displacement of the n th particle,

$$y_n = A \cos 2\pi(\nu t - ax) = A \cos 2\pi(\nu t - and) \quad (2.3)$$

where ν is the frequency, a the wave number or reciprocal wave length, A an arbitrary constant, and t the time. Now in Eq. (2.3) we may replace a by

$$a' = a \pm \frac{m}{d} \quad m \text{ an integer} \quad (2.4)$$

without changing the value of the displacement. This means that ν must be a periodic function of a with period $1/d$.

Now the phase velocity V , with which the waves propagate, is given by

$$V = \frac{\nu}{a} \quad (2.5)$$

Therefore, if we draw a curve of $\nu = \nu(a)$ as a function of a , the phase velocity for a given wave length will be given by the slope of the line drawn from the origin to the point on the $\nu(a)$ curve corresponding to the given wave length. The function $\nu(a)$ may

¹ "Popular Lectures," vol. I, p. 185.

be calculated and turns out to be

$$\nu(a) = B |\sin \pi a d| \quad (2.6)$$

where B is a constant that is a function of the constants of the lattice. From Eq. (2.6) we see that

$$V = \left| \frac{\nu(a)}{a} \right| = B \left| \frac{\sin \pi a d}{a} \right| = V_\infty \frac{|\sin \pi a d|}{|\pi a d|} \quad (2.7)$$

in agreement with Baden-Powell's equation (2.1), if we take

$$V_\infty = \pi d B \quad (2.8)$$

From Eq. (2.6) we see that $\nu(a)$ is a straight line for small values of a , *i.e.*, for large values of wave length. This means that the

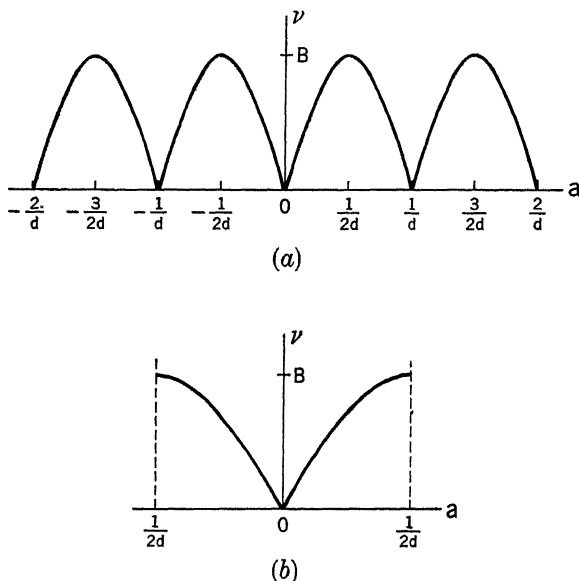


FIG. 2.4.—Frequency ν as a function of $a = 1/\lambda$ for the row of particles shown on Fig. 2.1.

velocity of propagation should be constant for large wave lengths, in agreement with the earlier calculations.

The curve of ν vs. a is shown in Fig. 2.4a. The periodicity of ν as a function of a means that for a given frequency the wave length is not completely determined. In fact, any a' , where a' is defined by Eq. (2.4), will give the same ν . The ambiguity in wave length results in an ambiguity in the direction of propaga-

tion—an uncertainty both in magnitude and in direction. This is easily seen by referring to Eq. (2.7).

The physical meaning of the ambiguity in wave length may be seen from Fig. 2.5. The solid circles give the equilibrium positions of the point masses and the open circles the displaced positions at some instant. Through the displaced positions are drawn three possible sine waves. All three waves give equally good descriptions of the motion, as far as observation of the

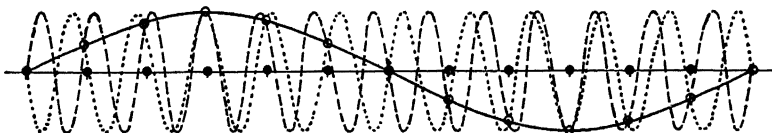


FIG. 2.5.—Different sine curves passing through the position of the particles.

points is concerned. The solid line gives the wave form for the only value of a such that

$$-\frac{1}{2d} \leq a \leq \frac{1}{2d} \quad (2.9)$$

Changing a by $1/d$ will take a out of this interval, as is immediately obvious. The dashed curve corresponds to $a + (1/d)$, and the dotted curve to $a - (1/d)$. A glance at the diagram shows that the solid and the dashed curves must propagate in the same direction for a given motion of the particles, and the dotted curve propagates in the opposite direction.

From now on, we shall adopt the convention expressed by Eq. (2.9). All ambiguity in wave length and direction of motion is removed if we restrict a to this interval, except in the two special cases where

$$a = \pm \frac{1}{2d} \quad (2.10)$$

We shall discuss these special cases shortly. The convention is not so arbitrary as might appear at first sight. It allows any wave length such that

$$\infty \geq \lambda \geq 2d \quad (2.11)$$

to have either direction of propagation, and excludes only wave lengths that lie in the interval

$$0 \leq \lambda \leq 2d \quad (2.12)$$

If we had a continuous structure so that the motion of all points lying on a straight line could be observed, the wave lengths included in the interval (2.11) would be the only ones observed, since in this case $d = 0$. Thus there will be no inconsistency in what we mean by wave length when we go from a continuous to a

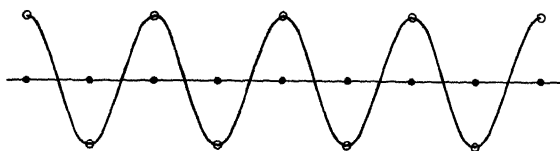
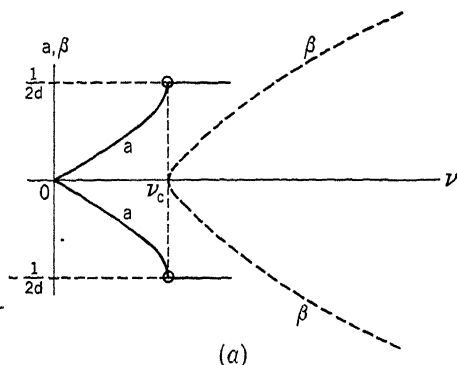
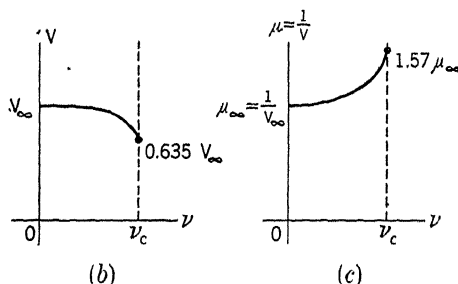


FIG. 2.6.—The limit $\lambda = 2d$.



(a)



(b)

(c)

FIG. 2.7.

discontinuous structure, and vice versa. Furthermore, the allowed interval contains a complete period of $\nu(a)$, so that none of the frequencies that can be propagated are omitted.

The special case noted in Eq. (2.10) is shown in Fig. 2.6. Here there is no way of distinguishing between the two possible wave numbers allowed by our convention, or between the two

possible directions of propagation. In fact, the wave might even be considered as a standing wave, *i.e.*, a superposition of the two allowed wave numbers. The wave length is, of course, in both cases $2d$.

Engineers frequently find it convenient to use other curves giving essentially the same information as our ν vs. a curve. The one of greatest interest is the attenuation curve (Fig. 2.7a). The solid part of the curve is our ν vs. a curve rotated through 90° .

The dotted, or β , part gives the attenuation β for frequencies higher than those that may be propagated. The attenuation will be discussed in detail in a later chapter. A lattice such as this, which allows propagation of all frequencies up to a maxi-

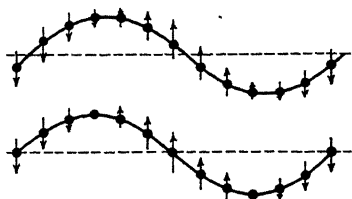


FIG. 2.8.—An example given by Lord Kelvin.

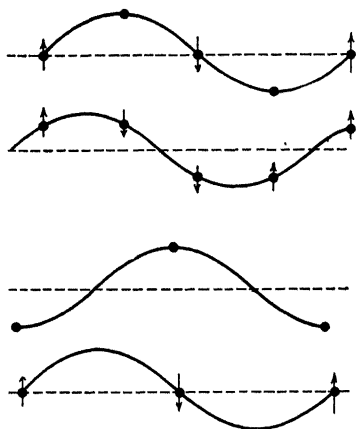


FIG. 2.9.—Other examples given by Lord Kelvin. The lower vibration corresponds to the limit $\lambda = 2d$.

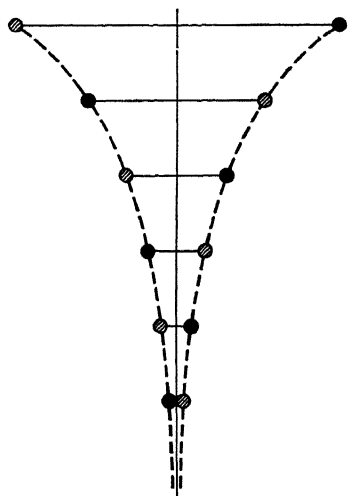


FIG. 2.10.—Attenuation of the wave for a frequency above cutoff (Lord Kelvin).

mum, or critical, frequency ν_c and damps all others, is called a *low-pass filter*; *i.e.*, it will pass low frequencies and stop higher frequencies. Figures 2.7b and 2.7c give V , the phase velocity, and μ , the index of refraction or reciprocal of phase velocity, as functions of frequency. Both curves terminate at the critical

frequency, as phase velocity is not defined for attenuated waves. The curves shown in Figs. 2.7*a*, *b*, and *c* are very useful for some practical purposes. However, in general, we shall find the ν vs. α curve (Fig. 2.4*b*) most useful for our analytical discussions.

Lord Kelvin's discussion is of great significance, since it contains the discovery of the cutoff frequency. Figures 2.8 to 2.11 are reproductions of Kelvin's original drawings and show the variation of wave velocity as a function of $N = \lambda/d$, the number of atoms per wave length. Modes of vibration are shown for large N and for $N = 2$ (cutoff), together with the attenuated wave corresponding to a frequency above cutoff. All this shows how clearly Kelvin understood the problem.

N	W
2	63,64
4	90,03
8	97,45
12	98,86
16	99,36
20	99,59
∞	100,00

FIG. 2.11.

The paper was often overlooked, since its title, "The Size of Atoms," did not imply any discussion of wave propagation. The connection is found in Cauchy's theory of dispersion. The curve in Fig. 2.2 shows that a material change in the wave velocity can be expected only if the wave length is just larger than $2d$. Hence, Cauchy's theory leads to the conclusion that interatomic distances should be just smaller than $\lambda/2$, giving a distance d of about 2,000 angstroms. This, however, sounded impossible since there was, at Kelvin's time, plenty of experimental evidence that interatomic distances could not amount to more than a few angstrom units. The thickness of oil films on water, for instance, had been measured and was quite well known.

Kelvin's conception of the molecular structure of matter may be illustrated by the following quotation:

I believe that by imagining each molecule to be loaded in a certain definite way by elastic connection with heavier matter . . . we shall

have a rude mechanical explanation for refractive dispersion. . . . It is not seventeen hours since I saw the possibility of this explanation.¹

This was a remarkable guess, which led Kelvin to the discovery of the modern refraction formula, usually known as the *Lorentz formula*.

3. Later Work on Models Similar to That Treated in Sec. 2

After analyzing Baden-Powell's work and discussing the critical wave length and frequency, Kelvin proceeded to devise a theory

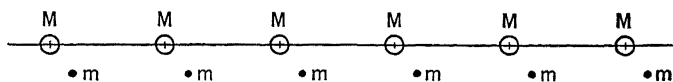


FIG. 3.1.—Kelvin's model for optical dispersion.

of dispersion based on a more complicated lattice than Baden-Powell's. He used the lattice shown in Fig. 3.1. Each of the masses in this model is supposed to have a small mass associated with it. The large masses are taken to have mass M and are the large circles in Fig. 3.1, while the small masses have mass m and are represented by dots. Each of the large masses interacts with the nearest large masses and with the small mass associated with it,

so that there are two elastic constants in the system. Introducing two masses effectively doubles the number of degrees of freedom of the system, and hence one would expect to find twice as many proper vibrations as if there were only one mass. The curve of ν vs. a is shown in Fig. 3.2. The curve is restricted to values of a between $\pm 1/2d$. It is seen that for each a there are two modes of vibration of the system, so that

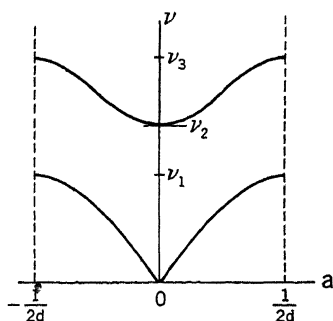


FIG. 3.2.

we do indeed have twice the number of modes obtained by Baden-Powell for his model with one mass. Frequencies below ν_1 and between ν_2 and ν_3 are propagated by the lattice, and all others are stopped. This lattice is an example of a *band-pass filter*. The interval between ν_1 and ν_2 is known as a *stopping band*, while that between ν_2 and ν_3 is known as a *passing band*. The frequencies ν_1 and ν_2 are very near the proper frequency of oscillation of one iso-

¹ *Op. cit.*, p. 194.

lated M - m molecule. This resonance frequency has nothing to do with the distance between molecules, and a material change in wave velocity is obtained when the resonance frequency lies in the near ultraviolet, just above the optical spectrum. Thus Kelvin explains refraction and escapes Cauchy's paradox.

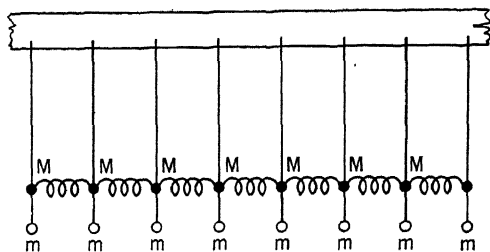


FIG. 3.3.—Vincent's model of the first mechanical filter.

Vincent¹ built a mechanical model to which Kelvin's theory was assumed to apply. The model is shown in Fig. 3.3. The large masses M are suspended from a beam on strings of equal length and connected to one another by springs. The small

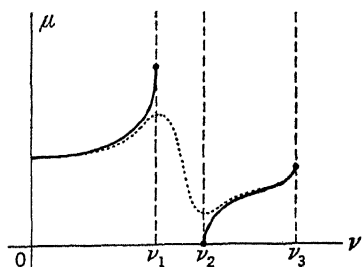


FIG. 3.4.—Index of refraction μ as a function of frequency ν for Vincent's model.

masses m are each suspended from one of the large masses. This model is evidently equivalent to Kelvin's more abstract scheme and was the *first mechanical filter* to be built. The motion of the system was observed for different frequencies. Vincent plotted curves of index of refraction μ against the frequency for comparison with standard dispersion curves. These curves are shown in Fig. 3.4. The solid curve is for negligible damping and the dotted curve for large damping. It is to be noted that the dotted curve is a typical anomalous dispersion curve. Vincent's curve of ν vs. a agreed with Kelvin's curve. The ratio $V = \nu/a$ can be measured on Fig. 3.2 and curve 3.4 obtained for $\mu = 1/V$ as a function of frequency ν .

Kelvin's paper received little notice, and the analogy between the propagation of electromagnetic radiation and the propagation of elastic waves along a loaded string was forgotten.

¹ *Phil. Mag.*, **46**, 537 (1898).

In 1887, Heaviside noted that increasing the inductance per unit length of a cable should reduce the attenuation of waves propagating along the cable. However, he discussed no experimental details. Two years later, in 1889, Vaschy tried loading a very long cable with four inductances, an experiment much too crude to give any observable result. In 1900, Pupin developed the analogy between mechanical and electric lines and, referring to Lagrange's work on the discontinuous string, succeeded in building loaded lines and low-pass electric filters. The line is shown in Fig. 3.5*a*. The inductances L' were spaced so that

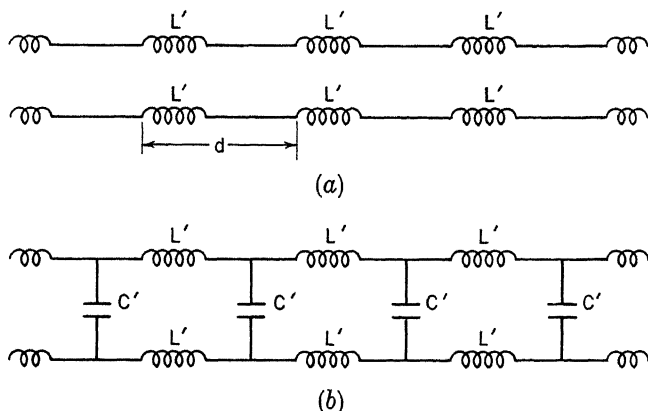


FIG. 3.5.—Low-pass electric filter and loaded line.

there were about ten inductances per wave length. Calling the capacitance per section d between the two halves of the line C' , Pupin obtained a critical frequency of

$$\nu_m = \frac{1}{\pi \sqrt{2L'C'}} \quad (3.1)$$

Figure 3.5*b* shows an equivalent line with the capacitance of the line lumped and placed along the line as indicated.

The first high-pass electric filter (*i.e.*, a line passing all frequencies higher than a certain critical frequency and stopping all others) was built by Campbell in 1906. The line is shown in Fig. 3.6. Campbell followed up his high-pass filter by designing various band-pass filters. Figure 3.7 is the band-pass filter analogous to Vincent's mechanical band-pass filter.

It is somewhat easier than in the analogous mechanical lines

to see why the electric lines mentioned above should pass some frequencies and stop others. The impedance offered by an electric circuit to a current passing through it is proportional to νL and inversely proportional to νC where ν is the frequency, L is the inductance, and C is the capacitance. Thus in the low-pass filter shown in Fig. 3.5b the impedance offered by the coils L' increases with the frequency, while the impedance of the capacities

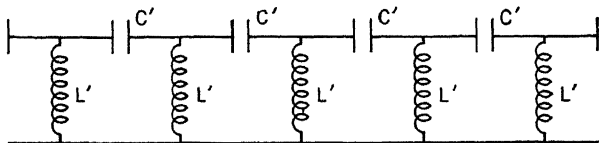


FIG. 3.6.—High-pass filter.

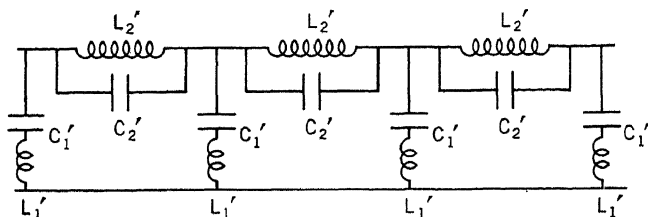


FIG. 3.7.—Band filter.

connected across the line decreases. The occurrence of a critical frequency is a result of the spacing and lumping of the inductances and capacities. In the high-pass filter the low frequencies will be shunted to the returning line through the inductances while the high frequencies will be passed. Again, the occurrence of a critical frequency is due to the discontinuous nature of the structure. These problems will be discussed in detail in a later section.

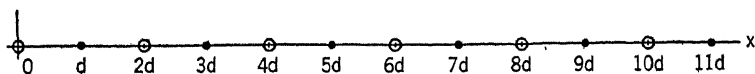


FIG. 3.8.—Born's model for sodium chloride.

In 1912, Born investigated the propagation of waves in crystals and rediscovered Kelvin's analysis. Using the model shown in Fig. 3.8, with large masses M and small masses m alternating at the points along the x axis defined by nd , where d is the distance between nearest neighbors, he obtained the curves shown in Figs. 3.9a and b. Figure 3.9a shows ν as a function of a . There are two branches to the curve because we have effectively

doubled the number of degrees of freedom of the system by adding another constant. The additional constant is, of course, the second value for mass. We shall find that in general the number of branches will equal the number of different masses occurring in the model; *i.e.*, the number of frequencies corresponding to a given wave number is equal to the number of degrees of freedom associated with each element or cell of the lattice. In this case the cell consists of a large mass and either of its neighboring small masses. If there were two different masses between

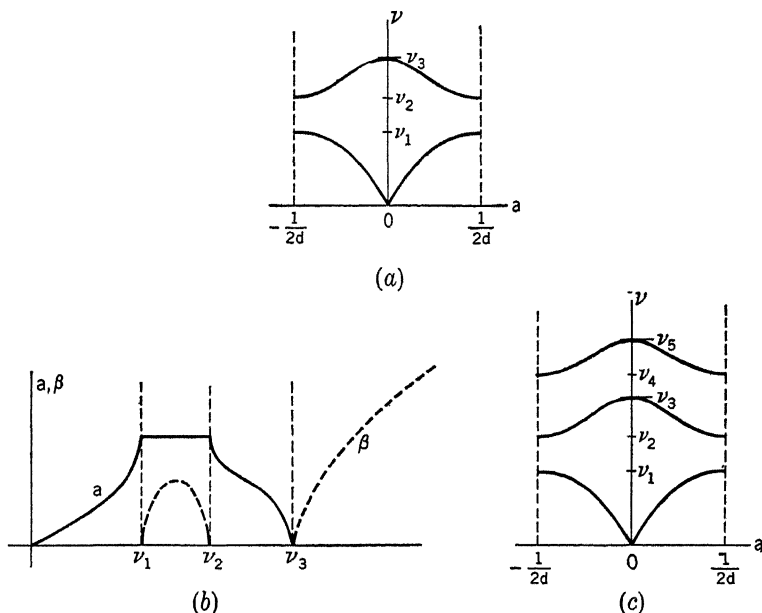


FIG. 3.9.

a given mass and the next one like it, and if this structure were repeated all along the lattice, each cell would have three degrees of freedom, and the ν vs. a curve would have one lower branch and two upper branches as in Fig. 3.9c. This property of discontinuous media will be discussed in greater detail later.

In general, the lower branch is called the *acoustical* branch. It corresponds to motion of the particles such that in each short section of the line all particles move in the same direction at a given instant. The upper branches are called *optical* branches and correspond to one or more types of particles moving in the

direction opposite to that of the others at any given instant. In Born's model, where we have only two types of particle, the optical branch corresponds to the motion of the large masses in one direction while the small masses move in the other.

Figure 3.9b is the attenuation curve for Born's model and represents the generalization of Fig. 2.7a. There are one stopping band and two passing bands associated with this model. The

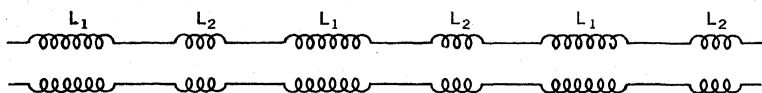


FIG. 3.10.—Electric filter corresponding to Born's sodium-chloride model.

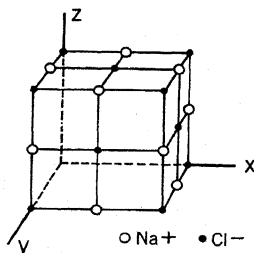


FIG. 3.11.—NaCl crystal lattice.

electrical analogue to Born's lattice is a line with small and large inductances alternating (Fig. 3.10).

Born's problem is usually referred to as the NaCl crystal lattice problem, since a very similar situation is found in the NaCl crystal structure: it is a cubic lattice with Na^+ and Cl^- ions alternately located at the lattice points, as shown in Fig. 3.11. Along one axis, the x axis, for instance, the structure is exactly the same as that in Fig. 3.8.

CHAPTER II

PROPAGATION OF WAVES ALONG ONE-DIMENSIONAL LATTICES.

GENERAL RESULTS AND QUALITATIVE DISCUSSION

4. General Remarks

Before proceeding to the mathematical treatment of waves propagating along a one-dimensional lattice, we shall make some general remarks about the problem and discuss some particular cases qualitatively. The simplest example of a one-dimensional lattice is Baden-Powell's model with equal masses spaced uniformly in a line. If we take the masses along the x axis, the x coordinate of the n th mass will be given by

$$x = nd + \psi_n \quad (4.1)$$

where ψ_n is the displacement of the n th particle from its equilibrium position. ψ_n may be taken to represent transverse or longitudinal displacement, or any other quantity whose value may be defined at the points occupied by the masses but not elsewhere (electric polarization, for instance); *i.e.*, we may regard ψ_n as a property associated with point mass n . This property is propagated as a wave if the physical problem admits a solution of the type

$$\left. \begin{aligned} \psi_n &= A e^{2\pi i(\nu t - a n d)} = A e^{i(\omega t - k n)} \\ a &= \frac{1}{\lambda}, \quad k = 2\pi a d, \quad \omega = 2\pi \nu \end{aligned} \right\} \quad (4.2)$$

where ν is the frequency, t the time, a the wave number, λ the wave length, d the period of the lattice, ω the angular frequency, k the product of the wave number and the period of the lattice multiplied by 2π , and A a constant amplitude. The quantity k is the change in phase in passing from a point n to its right-hand neighbor $n + 1$:

$$\psi_{n+1} = \psi_n e^{-ik} \quad (4.3)$$

Thus k is essentially defined as an angle and can be known only as *modulus* 2π . The same solution of the problem is obtained for

$$k \quad \text{or} \quad k' = k + 2m\pi \quad (4.4)$$

when m is a positive or negative integer. Equations of the physical problem must yield the same value of ω or ν for every equivalent k or k' , which means that the frequency ν is a periodic function of k or a :

$$\left. \begin{array}{ll} \omega = f(k) & \text{period } 2\pi \text{ in } k = 2\pi ad \\ \nu = F(a) & \text{period } \frac{1}{d} \text{ in } a \end{array} \right\} \quad (4.5)$$

This is a general and direct consequence of the periodic and discontinuous structure of the one-dimensional line. It was explained in Chap. I in Eq. (2.4) by saying that if ψ could be measured between particles, the uncertainty in k or a would be eliminated, but since ψ is measured only at the discrete points nd , the condition (4.5) is unavoidable.

On account of the periodic properties of the line, it is sufficient to discuss the properties of the functions f or F inside one period of k or a . The most convenient choice is

$$\left. \begin{array}{l} -\pi \leq k \leq \pi \\ -\frac{1}{2d} \leq a \leq \frac{1}{2d} \end{array} \right\} \quad (4.6)$$

since a wave always propagates in the same way to the right and to the left. This means that the functions f and F have the additional property of being *even* functions. Positive k means a wave propagating to the right; negative k a wave propagating to the left. If k_0 is a positive number in the fundamental interval (4.6), it represents a wave going to the right, and so does $k_0 + 2\pi$; but $k_0 - 2\pi$ is negative and represents a wave going to the left (Fig. 2.5). Hence, the uncertainty is not only in the magnitude of a or k but also in the direction of propagation.

The limitation (4.6) means

$$\lambda = \frac{1}{|a|} \geq 2d \quad (4.7)$$

The shortest wave length is thus equal to twice the distance between particles and corresponds to a certain *critical frequency*

or *cutoff frequency* ν_m that is characteristic of the structure. In many important cases ν_m is the maximum frequency, and the system works as a *low-pass filter* for all frequencies

$$\nu \leq \nu_m \quad (4.8)$$

Frequencies above ν_m are strongly attenuated. Condition (4.8) is, however, not the only possible one, and other situations may arise when ν_m would be a minimum. The system as a whole is always a filter, but it can be of the low-pass, high-pass, or band-pass type.

These general results, plus a direct discussion of the waves corresponding to the limiting cases, $\lambda = \infty$, $a = 0$, and $\lambda = 2d$, a maximum, may in a number of instances give enough information to enable one to describe, at least qualitatively, the general properties of the structure. In the next few sections we shall apply this discussion to specific examples of one-dimensional lattices.

5. A Lattice of Free Particles

By a lattice of free particles we mean particles in a one-dimensional lattice with no forces present except those due to interactions of the particles among themselves. For purposes of this discussion we shall limit the interactions to nearest neighbors. An example of this is a loaded elastic cord with the masses distributed uniformly, where the elasticity of the cord remains constant along its length and plays the part of the interaction forces.

Let us first consider longitudinal displacements. The case $a = 0$ corresponds to infinite wave length. In this case the lattice as a whole is displaced, and no change in the distance between masses occurs. Thus no force is brought into play. The frequency is zero. For $a \neq 0$, but still very small, the wave length is large compared with the distance between masses, and hence the waves are propagated as if the lattice were a continuous string. The velocity of propagation of waves along a continuous string is constant for all wave lengths; *i.e.*, for long wave lengths, the frequency is proportional to $|a|$. A rigorous treatment shows that the velocity decreases for wave lengths comparable with the distance between masses. Now if a wave is to be propagated at all, the frequency must be a periodic function of a . Further-

more, the curve of ν vs. a must be symmetrical about the origin. If it were not, the frequencies for a given wave length propagating in opposite directions would be different, a fact that would be in contradiction with the symmetry of the structure. If ν is to be both periodic and symmetrical about the origin, there must be a maximum in the value of ν at $1/2d$, since the period of ν is $1/d$. Thus we obtain a curve of the general shape of that in Fig. 2.4a. We shall, of course, justify the exact shape mathematically in a later section.

The remarks made on the longitudinal vibrations also apply to transverse vibrations. Qualitatively, they may be treated in just the same way. Quantitatively, however, there is a difference. The velocities of propagation for large wave length are

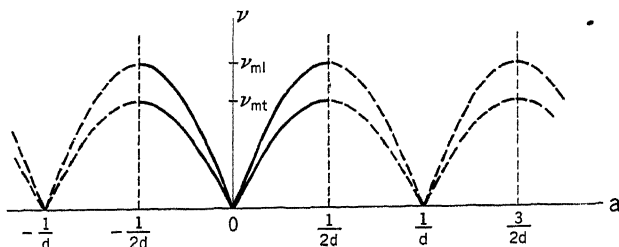


FIG. 5.1.—Longitudinal and transverse vibration along the row of particles shown on Fig. 2.1.

different in the longitudinal and transverse cases, and the maximum frequencies are also different. A typical curve for a one-dimensional lattice with particles with two degrees of freedom is shown in Fig. 5.1. The subscripts t and l on the maximum frequencies refer to transverse and longitudinal vibrations, respectively. The lower curve, representing transverse vibrations, should properly be considered a superposition of two branches of the same frequency, since there are two independent directions perpendicular to the lattice in which the masses might move. If there were an asymmetry in the elastic cord (*e.g.*, if it were of elliptical cross section), the lower branch would split into two distinct branches to give the extra frequencies demanded by the added degree of freedom. The solid curve corresponds to the interval (4.6), and its periodic continuation is shown as a dashed curve.

The transverse branches will usually be below the longitudinal branch in a loaded string, since the force required for a given dis-

placement is smaller in the transverse than in the longitudinal direction. The frequency of displacement is proportional to the square root of the elastic constant, which will be smaller in the case of transverse displacements.

6. Longitudinal Vibration in a Row of Equidistant Coupled Oscillators

A particle attracted to some equilibrium position by an elastic restoring force acts as a harmonic oscillator. It has one proper frequency ν_0 that depends on the elastic restoring force and the mass of the particle. If its elastic restoring force is different in the x , y , and z directions, we have what is called an *anisotropic oscillator*. An anisotropic oscillator has three proper frequencies,

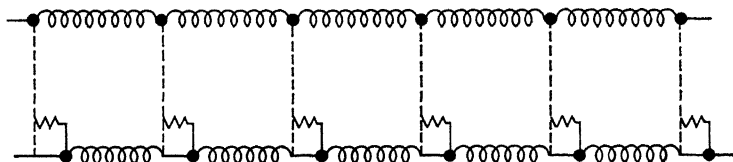


FIG. 6.1.—A row of harmonic oscillators coupled together.

ν_{0x} , ν_{0y} , and ν_{0z} , corresponding to vibrations in the x , y , and z directions, respectively.

Let us consider a row of similar harmonic oscillators (isotropic) spaced at distance d from one another along the x axis and allow interactions between nearest neighbors (Fig. 6.1). We wish to study the longitudinal modes of vibration of this system. For infinite wave length, $a = 0$. Infinite wave length means that all the particles are displaced simultaneously by the same amount. Since the distances between the particles do not change, the forces of interaction do not enter into the problem. Each particle is attracted to its equilibrium position with the same elastic force, and the system will oscillate with frequency ν_0 . For a slightly smaller wave length the particles will be displaced relatively to one another, and the forces of interaction will play a part in the motion of the system. The frequency associated with this wave length will be slightly different from ν_0 . Whether the frequency increases or decreases will depend on whether the resulting forces (elastic plus interaction) are larger or smaller than the restoring force tending to return each particle to its equilibrium position.

It may be shown that for large wave lengths ν is given by

$$\nu = \nu_0 + ba^2 \quad (6.1)$$

The sign of b depends on the constants of the system and determines whether ν shall increase or decrease as $|a|$ increases. As the wave length becomes comparable with $2d$, the considerations of the previous sections on one-dimensional lattices apply, and ν approaches an extremum. Thus we will have two limiting frequencies, ν_0 and ν_m (where m stands for maximum or minimum as the case may be). Frequencies between ν_0 and ν_m will propagate along the system, and other frequencies will be damped out. The system therefore forms a band-pass filter. The solid curve

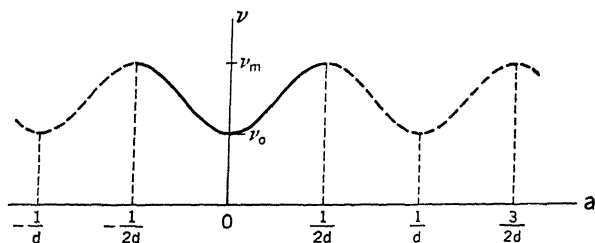


FIG. 6.2.—Frequency ν as a function of $a = 1/\lambda$ for the row of harmonic oscillators.

in Fig. 6.2 shows the curve ν vs. a in the fundamental interval (4.6) for the case $b > 0$. If each particle represented an anisotropic oscillator instead of an isotropic oscillator, there would be three curves, one for longitudinal and two for transverse vibrations. These curves might overlap and would not necessarily all rise as $|a|$ increases from zero.

7. Longitudinal Vibrations in a Row of Diatomic Molecules

The scheme described in the last section is somewhat artificial. It is rather difficult to imagine a particle in nature being tied to an equilibrium position by a little spring. A more realistic picture is obtained by considering diatomic molecules. This is a more complicated problem, since we must introduce a second type of particle that may interact with the first type as well as with its own type.

A lattice of diatomic molecules is shown in Fig. 7.1. The open circles are to have mass M , and the dots are to have mass m . An isolated molecule will have a certain proper frequency of vibra-

tion that we call ν_0 . This frequency corresponds to an oscillation of the two masses along the x axis in opposite directions in such a way that their center of mass remains at rest.

Let us consider the motion of a row of diatomic molecules spaced at distance d from one another along the x axis. We assume, of course, that the molecules interact, but we limit the interaction to nearest neighbors. There will now be *two* wave functions, both imaginary exponential, one describing the motion of the masses M and the other describing the motion of the masses m . These two functions may be written

$$\psi_M = A_M e^{2\pi i(\nu t - ax)} \quad \text{and} \quad \psi_m = A_m e^{2\pi i(\nu t - ax)} \quad (7.1)$$

The frequencies and wave numbers will be the same, but the amplitudes may be different. The frequency ν may be found as

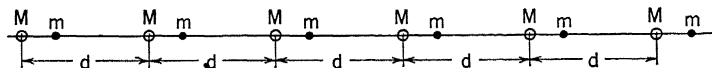


FIG. 7.1.—A row of diatomic molecules.

a function of the constants of the system and of a . It turns out to be double valued in ν , as will be shown in the rigorous theory, corresponding to the doubly infinite set of degrees of freedom of the system.

For infinite wave length, the atoms all oscillate in phase, and we may take

$$A_M = A_m \quad (7.2)$$

This corresponds to a translation of the lattice as a whole without alteration of the distance between particles, and hence the frequency is zero. Another frequency for infinite wave length is obtained if we take the small and large masses moving in opposite directions in such a way that the centers of gravity of the molecules remain at rest. This frequency would be ν_0 if there were no interaction between molecules. The presence of interactions would change this frequency. If the wave length is decreased, the lower branch of the ν vs. a curve will rise. This branch is just what would be obtained if we took each molecule to be a single particle. The upper branch will increase or decrease from its frequency at $a = 0$, depending on the relative values of the constants involved. Figure 7.2 shows the frequency curves. The limit to the frequency of the upper branch is ν'_0 for $a = 0$. Either, but not both, of the two upper branches shown may

occur. Figure 3.2 (Vincent) and Fig. 3.9 (Born) represent two typical examples with different upper curves. The size of ν_0 relative to the maximum frequency of the lower branch depends on the constants of the system, as does also the width of the

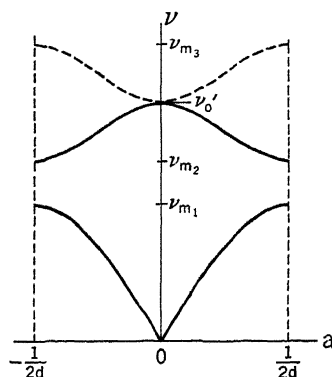


FIG. 7.2.—Frequency ν as a function of $a = 1/\lambda$ for a row of diatomic molecules.

upper branch. Frequencies located in the stopping bands may be shown to decay exponentially, as in the other models we have discussed. The a corresponding to these frequencies are complex with imaginary part β . β is therefore the attenuation constant for a given frequency. The attenuation curves are shown

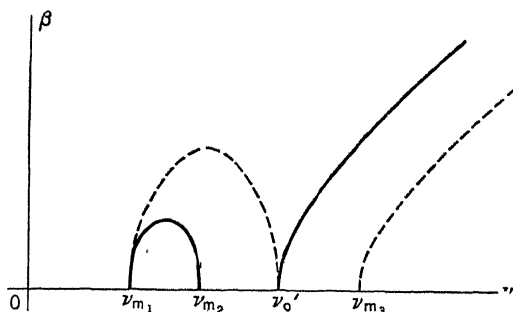


FIG. 7.3.—Attenuation as a function of frequency for a row of diatomic molecules.

in Fig. 7.3. The solid curve is for the solid upper branch and the dashed curve for the dashed upper branch of Fig. 7.2.

In these examples, the following features can be recognized that will be proved in the detailed analysis of later chapters:

1. Periodicity of ν as a function of k or a (4.5).

2. The possibility of a reduction of k or a inside the fundamental interval (4.6).

3. If the elementary cell of the one-dimensional lattice contains a system with N degrees of freedom, there will be N different waves corresponding to each k value, with N different frequencies. Examples with $N = 1, 2, 3$ were given in Secs. 5, 6, and 7.

4. Hence, the number of degrees of freedom inside an elementary cell equals the number of branches in the curve $\nu = F(a)$ and the number of passing bands of the structure (with possible overlapping of the passing bands).

5. Frequencies outside the passing bands are not propagated but decay exponentially along the line.

These are the general properties of one-dimensional periodic structures that will be investigated mathematically in the following sections.

CHAPTER III

MATHEMATICAL TREATMENT OF A ONE-DIMENSIONAL LATTICE OF IDENTICAL PARTICLES

8. Equation of Motion of a One-dimensional Lattice of Identical Particles

In this and the following sections we shall derive rigorously the results discussed qualitatively in the first two chapters. We shall assume an infinite lattice of identical particles of mass M . The particles in equilibrium are separated by a distance d along the x axis, and we shall take the oscillations of the particles to be longitudinal. We number the particles by calling the particle at the origin 0, the next particle to the right 1, etc. The displacement of the n th particle is denoted by y_n , so that x_n , the coordinate of particle n , will be given by

$$x_n = nd + y_n \quad (8.1)$$

We shall assume interactions between all particles, and for this we require the expression for the distance between two particles n and $n + m$. This distance is

$$r_{n,n+m} = x_{n+m} - x_n = md + y_{n+m} - y_n \quad (8.2)$$

This expression may be either positive or negative, depending on whether m is positive or negative. The energy of interaction between two particles will be expressed as a potential function that will be assumed to depend only on the distance between the two particles:

$$U(r) = U(|x_{n+m} - x_n|) \quad (8.3)$$

The total potential energy of the lattice will then be given by

$$U = \sum_n \sum_{m>0} U(|x_{n+m} - x_n|) \quad (8.4)$$

m must be restricted to positive values so that the interaction between a given pair of particles will be counted only once. We

might take the sum over all values of m and divide by two to compensate for counting each pair of particles twice. However, we prefer to restrict m to positive values, since this enables us to drop the absolute-value sign in the argument of U . If we assume that the displacements y_n are small compared with d , we may expand U in a Taylor series. Thus

$$U(x_{n+m} - x_n) = U(md) + (y_{n+m} - y_n)U'(md) \\ + \frac{1}{2}(y_{n+m} - y_n)^2U''(md) + \dots,$$

where $U'(md)$ and $U''(md)$ are the derivatives $\partial U/\partial r$ and $\partial^2 U/\partial r^2$ evaluated at md . Substituting the Taylor expansion in Eq. (8.4), and neglecting powers of $(y_{n+m} - y_n)$ higher than the second, we obtain for the potential energy of the lattice

$$U = \sum_n \sum_{m>0} \left[U(md) + (y_{n+m} - y_n)U'(md) \right. \\ \left. + \frac{1}{2}(y_{n+m} - y_n)^2U''(md) \right],$$

or

$$U = \text{const.} + \sum_n \sum_{m>0} \left[(y_{n+m} - y_n)U'(md) \right. \\ \left. + \frac{1}{2}(y_{n+m} - y_n)^2U''(md) \right], \quad (8.5)$$

where the constant is given by

$$\text{Const.} = \sum_n \sum_{m>0} U(md) = n \sum_{m>0} U(md)$$

The force F_p acting on the p th particle is obtained by taking the negative derivative of the potential energy with respect to the displacement of this particle. Before performing the differentiation it should be noted that only two terms from the sum over all values of n will remain, the others dropping out because they do not contain the variable y_p . The two remaining terms will be those for which $n = p$ and $n + m = p$. m is to be positive, so the terms for which $n = p$ will give the force on particle p due to particles to the right, while terms for which $n + m = p$ give the force on particle p due to particles on the left. Therefore,

$$\begin{aligned}
 F_p &= -\frac{\partial \mathcal{U}}{\partial y_p} = -\frac{\partial}{\partial y_p} \sum_n \sum_{m>0} \left[(y_{n+m} - y_n) U''(md) + \frac{1}{2} (y_{n+m} - y_n)^2 U'''(md) \right] \\
 &= -\frac{\partial}{\partial y_p} \sum_{m>0} \left[(y_{p+m} - y_p) U''(md) + \frac{1}{2} (y_{p+m} - y_p)^2 U'''(md) \right. \\
 &\quad \left. + (y_p - y_{p-m}) U''(md) + \frac{1}{2} (y_p - y_{p-m})^2 U'''(md) \right] \\
 &= -\sum_{m>0} [-U''(md) - (y_{p+m} - y_p) U'''(md) \\
 &\quad + U''(md) + (y_p - y_{p-m}) U'''(md)] \quad (8.6a)
 \end{aligned}$$

or, writing U''_m instead of $U''(md)$,

$$F_p = \sum_{m>0} U''_m (y_{p+m} + y_{p-m} - 2y_p) \quad (8.6b)$$

These formulas require some discussion and explanation. In Eq. (8.6a), for instance, we find in the first row a term $-U''(md)$ representing the force of atom $(p+m)$ on atom p . In an

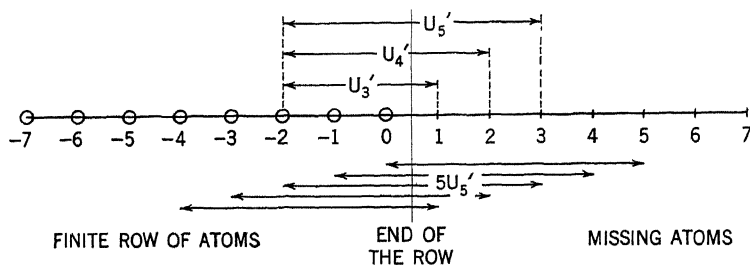


FIG. 8.1.

infinite lattice this term is compensated by an opposite force $+U''(md)$ found in the second row of Eq. (8.6a) and representing the force of atom $(p-m)$ on atom p .

The situation is different in a *finite* lattice (Fig. 8.1). Let us assume the row of atoms to extend from $n = -\infty$ to $n = 0$, with all atoms $n = 1, 2, 3, \dots$ missing, and let us discuss the forces to be added in order to keep the structure undisturbed near the end of the row. External forces that would make up exactly for the forces that the missing atoms would produce on

the end of the row must be provided for. For instance, we must add the forces

$$U'_3, U'_4, U'_5, \dots$$

on atom $n = -2$. This means a very complicated set of forces acting on the last atoms of the row, if the row is to be kept unperturbed with the constant distance d up to the last atom. The *total force* required on all the last atoms of the row is

$$F_t = U'_1 + 2U'_2 + 3U'_3 + \dots = \sum_{m=1}^{\infty} mU'_m \quad (8.7)$$

since there are m pairs of atoms interacting at distances md across the border. The sketch in Fig. 8.1 visualizes the situation for $m = 5$. In order to obtain a one-dimensional lattice with distance d between neighboring particles, it is *necessary* that the *total force* acting upon the end of the lattice be F_t , but the condition is not sufficient.

If this total force F_t is differently distributed between the particles at the end of the row, two things may happen:

1. It is possible that a local perturbation of the row is produced near the end, but that at large distances from the end the equidistance d is obtained. This is usually the case, with forces decreasing rapidly when the distance is increased, such as the ones encountered in most physical problems of crystal lattices. If the forces extend only to a distance Ld , the sum in Eq. (8.7) must be taken from $m = 1$ to $m = L$, and the distance upon which the perturbation of the lattice occurs is of the order of Ld .

2. The perturbation may extend throughout the lattice and offer a periodic character as a function of the distance, thus resulting in a sort of superlattice or periodic structure with a distance $D > d$. There may also be different values d_1, d_2, \dots corresponding to the same total end force F_t .

For instance, a free row of particles is one terminating freely with no external forces added. This means that no perturbation will occur only if all terms $U'_1 = U'_2 = \dots = U'_L = 0$, and in this case the lattice will keep the interval d up to its end. If all U'_m are not zero, a perturbation appears near the end of the lattice (case 1) or even along the whole lattice (case 2).

This one-dimensional example corresponds to the problem of

surface structure and *surface tension* for solids or liquids. In the three-dimensional problems of physics, the interaction between particles decreases very rapidly for increasing distances, and case 1 above is practically always obtained. The last L atoms of each row build a surface layer Ld deep, which surrounds the solid or liquid structure. The perturbation of the lattice inside this surface layer results in additional forces, the resultant of which is known as surface tension.

The type of perturbation in the lattice and the extent of this perturbation will be discussed later on (see Sec. 10), but we should immediately emphasize the *great complexity of the boundary conditions* for structures including *particles interacting at large distances*. The situation at the boundary cannot be defined by a set of forces acting on the last particles, but the whole distribution of these forces on the different particles at the end of the row must be specified. The usual mathematical statements about forces on the boundary are completely inadequate. A similar situation will be found in connection with problems of wave propagation across the junction of two lattices, or reflection of waves at the boundary of a lattice (see Sec. 24), where a minute description of the type of junction extending all through a boundary layer of order of thickness Ld would be required.

As for Eq. (8.6b) and vibrations inside an infinite lattice, the force F_p will be balanced by the inertial force so that the equation of motion for the system will be

$$F_p = M \frac{d^2 y_p}{dt^2} = \sum_{m>0} U''_m (y_{p+m} + y_{p-m} - 2y_p) \quad (8.8)$$

Let us assume a wave solution for Eq. (8.8).

$$y_p = A e^{2\pi i(\nu t - a x_p)} = A e^{2\pi i(\nu t - a p d)} \quad a = \frac{1}{\lambda} \quad (8.9)$$

ν is, of course, the frequency and a the wave number. This gives

$$\begin{aligned} y_{p+m} + y_{p-m} - 2y_p &= A e^{2\pi i(\nu t - a p d)} (e^{-2\pi i m d a} + e^{2\pi i m d a} - 2) \\ &= -2y_p (1 - \cos 2\pi a m d) = -4y_p \sin^2 \pi a m d \end{aligned}$$

Therefore, Eq. (8.9) will be a solution of Eq. (8.8) if the following relation between ν and a is satisfied:

$$\begin{aligned} \frac{M\omega^2}{4} &= M\pi^2\nu^2 = \sum_{m>0} U''_m \sin^2 \pi amd \\ &= \frac{1}{2} \sum_{m>0} U''_m (1 - \cos 2\pi amd) \quad (8.10) \end{aligned}$$

with $U''_m = U''(md)$. From Eq. (8.10) we may verify at once that ν is a periodic function of a and has period $1/d$, since

$$\nu^2 \left(a + \frac{1}{d} \right) = \nu^2(a)$$

and ν must be positive.

9. Rigorous Discussion for the Case of Interactions between Nearest Neighbors Only

If we assume that the interactions among the particles are negligible except for nearest neighbors, Eq. (8.10) reduces to

$$M\pi^2\nu^2 = U'' \sin^2 \pi ad \quad U'' = U''_1 \quad (9.1)$$

This is the equation on which the qualitative discussions in the first two chapters were based. We may compute the velocity of propagation of the wave.

$$V = \frac{|\nu|}{a} = \sqrt{\frac{U'''}{M}} \frac{|\sin \pi ad|}{|\pi a|} = d \sqrt{\frac{U'''}{M}} \frac{|\sin \pi ad|}{|\pi ad|} \quad (9.2)$$

The velocity for infinite wave length V_∞ is therefore

$$V_\infty = d \sqrt{\frac{U'''}{M}} \quad \lambda \rightarrow \infty, \quad a \rightarrow 0 \quad (9.2a)$$

and Eq. (9.2) checks with Baden-Powell's equation (2.1).

In order to set up the connection between these results and Newton's calculation for the velocity of sound in air, we must define a modulus for our discontinuous system; and this must be done in such a way that in the limit of dense spacing of our particles (*i.e.*, a continuous structure) the modulus will go over into the ordinary extension modulus, defined as tension divided by strain. In our discontinuous structure, we can define the tension between two particles as simply the force between them, and this will be equal, for the p th and $(p+1)$ st particles, to

$$U''(d)(y_{p+1} - y_p)$$

since the resultant force on the p th particle, due to both particles $(p + 1)$ and $(p - 1)$, is

$$U'''(d)(y_{p+1} + y_{p-1} - 2y_p)$$

Furthermore, we can define the strain between particles p and $(p + 1)$ as $(y_{p+1} - y_p)/d$. The modulus will, accordingly, be

$$\epsilon = dU''(d) \quad (9.3)$$

and it is evident that in the limiting case of dense spacing all our definitions will go over into the usual definitions.

If we call our modulus ϵ and the average linear density of our system ρ (i.e., $\rho = M/d$), Eq. (9.2a) becomes

$$V_\infty = \sqrt{\frac{\epsilon}{\rho}} \quad (9.3a)$$

which is Newton's formula [Eq. (1.1)] with ϵ in place of Newton's bulk modulus ed . We can identify our U'' with Newton's elastic constant e .

For the wave length large compared with d , i.e., if the lattice may be regarded as a continuous medium, the velocity is V_∞ and is independent of the wave length. As the wave length decreases, the velocity decreases and approaches $2V_\infty/\pi$, or 0.635 times V_∞ , the value for infinite wave length (see Fig. 2.2). This velocity is reached at the wave length $\lambda = 2d$. For $\lambda = 2d$, there is an ambiguity in the velocity of propagation, as pointed out in an earlier section, since the wave may be propagating in either direction with velocity $0.635V_\infty$ or may be a standing wave. The cutoff frequency ν_m is obtained from Eq. (9.1) by setting $ad = \frac{1}{2}$.

$$\nu_m = \frac{1}{\pi} \sqrt{\frac{U''}{M}} \quad (9.4)$$

For frequencies lower than the limiting frequency ν_m we obtain real solutions for a . For higher frequencies a is complex, since

$$\pi^2 \nu^2 M = U'' \sin^2 \pi ad \quad (9.1)$$

If we set

$$\left. \begin{aligned} a &= \pm \frac{1}{2d} \pm i\beta, & k &= 2\pi ad = \pm\pi \pm i2\pi\beta d \\ \sin \pi ad &= \pm \sin \frac{\pi}{2} \cos i\pi\beta d = \pm \cosh \pi\beta d \end{aligned} \right\} \quad (9.5)$$

then

$$\pi^2 \nu^2 M = U''' \cosh^2 \pi \beta d \quad (9.6)$$

or

$$|\nu| = \frac{1}{\pi} \sqrt{\frac{U'''}{M}} |\cosh \pi \beta d| \quad (9.7)$$

β is called the *attenuation coefficient*, and in the attenuation curves the magnitude of β is plotted as a function of ν . Curves representing the real and imaginary parts of $a = \alpha + i\beta$ as functions of the frequency ν have been drawn in Fig. 2.7 (Sec. 2). Between 0 and ν_m , a is real, and above ν_m the real part of a keeps a constant value $\pm 1/2d$ while the imaginary part β increases very rapidly. This means that for frequencies above the cutoff ν_m the vibration decays exponentially along the string (β term) while successive atoms oscillate in opposite directions (real part $1/2d$). This is easily seen in Fig. 2.10, which is a reproduction of one of Kelvin's original drawings. It shows that Kelvin had actually grasped all the details of this problem.

10. Discussion of the Distance of Interaction

In the case of interactions between nearest neighbors only, we find that there is a single frequency corresponding to a given wave length and that there is only one wave length larger than $2d$ for each frequency. Now if the interactions extend to the L th neighbor, *i.e.*, to a distance of Ld , we obtain the following expression relating frequency and wave number [Eq. (8.10)]:

$$\begin{aligned} \pi^2 \nu^2 M &= \sum_{0 < m < L} U'''_m \sin^2 \pi a m d \\ &= \frac{1}{2} \sum_{0 < m < L} U'''_m (1 - \cos 2\pi a m d) \quad (10.1) \end{aligned}$$

For very large wave lengths

$$V_\infty^2 = \frac{\nu^2}{a^2} = \frac{1}{\pi^2 M} \sum_{0 < m < L} U'''_m \frac{\sin^2 \pi a m d}{a^2} = \frac{d^2}{M} \sum_{0 < m < L} U'''_m m^2 \quad (10.2)$$

Thus V_∞ is still a constant whose value depends on the constants of the system. As the wave length decreases, the velocity of

propagation varies. The frequency corresponding to the limiting wave length $\lambda = 2d$, $a = 1/2d$, may be computed.

$$\nu^2 = \frac{1}{\pi^2 M} \sum_{0 < m < L} U''_m \sin^2 \frac{\pi m}{2} = \frac{1}{\pi^2 M} \sum_{\substack{0 < m < L \\ m \text{ odd}}} U''(md) \quad (10.3)$$

since

$$\sin \frac{\pi m}{2} = \begin{cases} 0 & m \text{ even} \\ \pm 1 & m \text{ odd} \end{cases} \quad (10.4)$$

so that the even terms in the sum drop out.

Returning to the general equation for ν [Eq. (10.1)], we note that to each value of a there will correspond a single frequency regardless of the extent of the interactions. Now $\cos 2\pi a m d$ may be expanded as a polynomial of degree m in $\cos 2\pi a d$. Thus

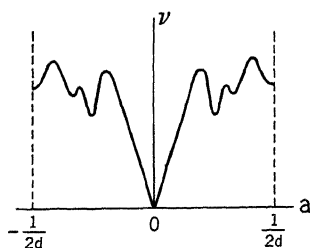


FIG. 10.1.

the frequency will be expressed as a polynomial of degree L in $\cos 2\pi a d$. This means that for a given frequency there will be L solutions for $\cos 2\pi a d$ and hence L solutions for a in the interval $-1/2d$ to $+1/2d$. The result of these remarks is that ν is a single-valued function of a , but a is not a single-valued function of ν , as shown in Fig. 10.1. It is not neces-

sary in this case that the maximum value of the frequency appear at the ends of the interval $-1/2d \leq a \leq +1/2d$, but the curve must end with a horizontal tangent in any case.

The L solutions for a for a given frequency need not all be real; some may be imaginary or complex. Such solutions are to be interpreted as meaning that the wave decays exponentially along the lattice. This is of special importance in the case of a finite lattice such as the one already discussed in Sec. 8 with Fig. 8.1. If we assume a sinusoidal motion of frequency ν imposed on the last particle of the lattice, the different waves corresponding to this frequency will be excited in various proportions. Those for which a is real will propagate along the lattice, and those for which a is imaginary or complex will decay exponentially from the point of excitation. If we wish to excite only one of the waves on a semiinfinite row of particles, we must impose on the

first L particles the motion characteristic of this special wave. In the case of interactions between nearest neighbors only, the boundary conditions were simple: we had only to specify the motion of the first particle. However, added interactions complicate the procedure, and the boundary conditions must be specified over a length Ld of the lattice.

The problem of the lattice at rest corresponds to the case $\nu = 0$. In drawing the curve in Fig. 10.1, it was assumed that the forces between the particles were such as to give only one real solution a for low ν values. The remaining $(L - 1)$ solutions must then be complex and result in a perturbation of the lattice that would decay exponentially from the border. The whole distance over which these exponential perturbations extend (at the limit $\nu = 0$) represents the thickness of the border in the one-dimensional case or of the surface layer in the three-dimensional problem. This assumption corresponds to case 1 discussed in Sec. 8 after Eq. (8.7). Another possibility would correspond to a curve going down to $\nu = 0$ for some $\pm a_1$ value of a , such as the curve of Fig. 10.2. Under such circumstances a steady periodic perturbation of wave length $\lambda_1 = 1/a_1$ may obtain throughout the lattice and realize a superlattice structure of period

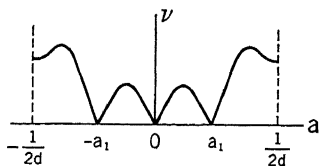


FIG. 10.2.

$$\lambda_1/d = 1/a_1d_1$$

as anticipated in Sec. 8, case 2.

Equation (10.1) gives ν^2 as a finite Fourier expansion in a . We may use Fourier's theorem to obtain the interactions among the various particles if we assume $\nu = F(a)$ is a known function.

$$U''(md) = -4\pi^2Md \int_{-1/2d}^{1/2d} F^2(a) (\cos 2\pi amd) da \quad (10.5)$$

As an example, let us seek the interactions that would give a constant velocity of propagation W throughout the passing band. Then

$$\nu = W|a|, \quad \nu^2 = F^2(a) = W^2a^2$$

a is, of course, to be taken in the usual interval. Curves corresponding to this problem are shown in Fig. (10.3). Then

$$\begin{aligned}
 U''_m &= -4\pi^2 M d W^2 \int_{-1/2d}^{1/2d} a^2 (\cos 2\pi a m d) da \\
 &= (-1)^{m-1} 2M \frac{W^2}{m^2 d^2} \quad (10.6)
 \end{aligned}$$

Now U''_m is the second derivative of the interaction energy of the two particles separated by md and appears as a function defined at discrete points at intervals of d along the x axis. We may take the continuous function

$$U'''(x) = -2 \frac{MW^2}{x^2} \cos \frac{\pi x}{d} \quad (10.7)$$

to represent the discontinuous function. The function (10.7) has the same values as $U'''(md)$ at the points where $U'''(md)$ is

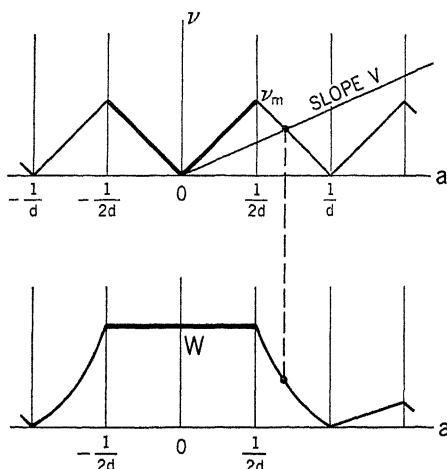


FIG. 10.3.

defined, but it is continuous, and hence we may integrate twice to find the interaction energy. The integration must be done by tables in this case. Once the function $U''(x)$ is known, however, one may construct a discontinuous line with the proper elastic forces between the elements to obtain a low-pass mechanical filter having a constant velocity of propagation for all frequencies in the passing band. The same method may be applied to a high-pass filter or to more complicated filters having one or more passing bands. For this simple example we may easily obtain ω^2 , where ω is the angular frequency, 2π times the frequency ν , as a Fourier series.

$$\begin{aligned}
 \omega^2 &= 4\pi^2\nu^2 = 4 \sum_m \frac{U''_m}{2M} (1 - \cos 2\pi amd) \\
 &= \sum_m \frac{4(-1)^{m-1}W^2}{m^2d^2} (1 - \cos 2\pi amd) \\
 &= \frac{4W^2}{d^2} \left[(1 - \cos 2\pi ad) - \frac{1}{4} (1 - \cos 4\pi ad) \right. \\
 &\quad \left. + \frac{1}{9} (1 - \cos 6\pi ad) - \frac{1}{16} (1 - \cos 8\pi ad) + \dots \right] \\
 &= \frac{4W^2}{d^2} \left[\left(1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \dots \right) - \cos 2\pi ad \right. \\
 &\quad \left. + \frac{1}{4} \cos 4\pi ad - \frac{1}{9} \cos 6\pi ad + \frac{1}{16} \cos 8\pi ad - \dots \right] \\
 &= \frac{4W^2}{d^2} \left[\frac{\pi^2}{12} - \cos 2\pi ad + \frac{1}{4} \cos 4\pi ad \right. \\
 &\quad \left. - \frac{1}{9} \cos 6\pi ad + \frac{1}{16} \cos 8\pi ad \dots \right] \quad (10.8)
 \end{aligned}$$

since

$$\frac{\pi^2}{12} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \dots \quad (10.9)$$

Let us replace $2\pi ad$ by k and recall that $a = \nu/W$ to obtain

$$\begin{aligned}
 k^2 &= 4\pi^2 a^2 d^2 = \frac{4\pi^2 \nu^2 d^2}{W^2} \\
 &= 4 \left(\frac{\pi^2}{12} - \cos k + \frac{1}{4} \cos 2k - \frac{1}{9} \cos 3k + \dots \right) \quad (10.10)
 \end{aligned}$$

Thus we have k^2 as a well-known Fourier expansion in k in the interval $-\pi, +\pi$.

11. The Low-pass Electric Filter

The electric filter shown in Fig. 11.1 is a low-pass electric filter. The equal self-inductances L alternate with equal capacities C . The capacities shunt out the high frequencies, and the low frequencies are allowed to pass. To obtain the equations of this line, we call Q_n and V_n the charge and potential, respectively, on condenser n , while i_n will be the current flowing between condensers $(n-1)$ and n . Then

$$L \frac{di_n}{dt} = V_{n-1} - V_n = \frac{Q_{n-1}}{C} - \frac{Q_n}{C}$$

and

$$i_n - i_{n+1} = \frac{dQ_n}{dt} \quad (11.1)$$

since

$$V_n = \frac{Q_n}{C} \quad (11.2)$$

Differentiating Eq. (11.1), we obtain

$$L \frac{d^2 i_n}{dt^2} = \frac{1}{C} \left(\frac{dQ_{n-1}}{dt} - \frac{dQ_n}{dt} \right) = \frac{1}{C} (i_{n-1} + i_{n+1} - 2i_n) \quad (11.3)$$

The solution of Eq. (11.3) gives the flow of current in the line, and from this the potential differences and charges on the condenser plates may be found. Equation (11.3) is identical with

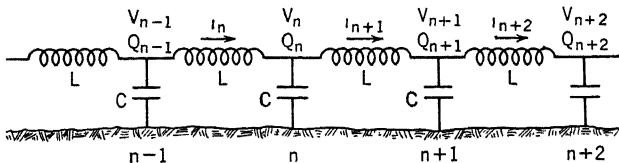


FIG. 11.1.

the equation of motion of a one-dimensional mechanical lattice [Eq. (8.8)] with interaction between nearest neighbors only (Chap. I, Sec. 2, or Chap. III, Sec. 9).

$$M \frac{d^2 y_n}{dt^2} = U''_1 (y_{n-1} + y_{n+1} - 2y_n) \quad (11.4)$$

U''/M is replaced by $1/LC$, and y_n is replaced by i_n . Thus all the results obtained for the low-pass mechanical filter apply automatically. The velocity of propagation for very long waves is d/\sqrt{LC} where d is the distance between condensers; there is a cutoff frequency ν_m , and all frequencies higher than ν_m decay exponentially; ν is a periodic function of the wave number. From Eq. (9.4) we may compute the cutoff frequency.

$$\nu_m = \frac{1}{\pi \sqrt{LC}} \quad (11.5)$$

The low-pass electric filter shown in Fig. 11.1, to which Eq. (11.3) applies, contains no resistance. Introduction of resistance changes the properties of the line slightly. There will be a slight attenuation of frequencies in the passing band due to energy losses in the resistance, and the cutoff frequency will be less

abrupt; *i.e.*, there will be a region of rapidly increasing attenuation for increasing frequency near ν_m . This problem will be discussed in detail in Chap IX. The curves in Fig. 2.7 will be changed into those in Fig. 11.2.

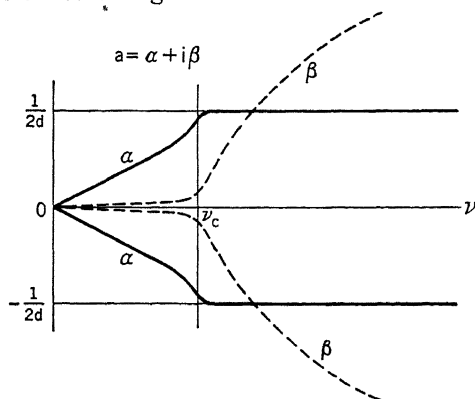


FIG. 11.2.

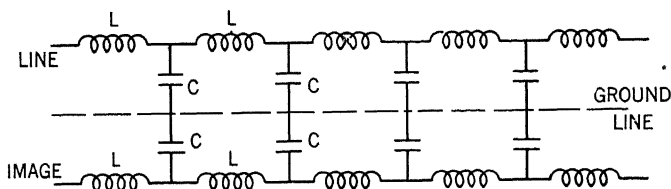


FIG. 11.3.

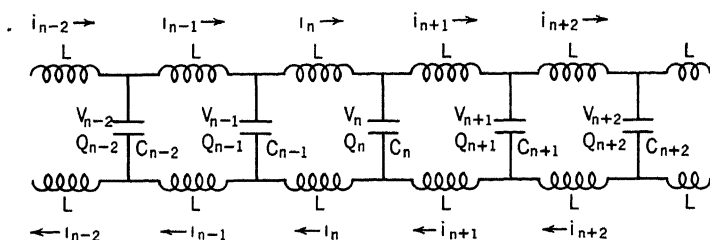


FIG. 11.4.

The single-line structure of Fig. 11.1 is equivalent to a double line (Fig. 11.3) constructed from the original line of Fig. 11.1 and its image. This can be simplified in the scheme of Fig. 11.4 with the same L values as in the single line but with capacities $\frac{1}{2}C$.

$$L' = L, \quad C' = \frac{C}{2} \quad (11.6)$$

Hence, the double line of Fig. 11.4 has exactly the same properties as the single line, with the values

$$\begin{aligned} v_{\infty} &= \frac{d}{\sqrt{LC}} = \frac{d}{\sqrt{2L'C'}} \\ v_m &= \frac{1}{\pi \sqrt{LC}} = \frac{1}{\pi \sqrt{2L'C'}} \end{aligned} \quad (11.7)$$

as announced in Eq. (3.1).

12. Analogies between Electrical and Mechanical Systems

In the last section we saw that the equation for the propagation of electric waves along a low-pass electric line was of exactly the same form as that for the propagation of elastic waves along a low-pass mechanical lattice. This suggests the possibility of making an analogy between electrical and mechanical lines that will hold generally. The detailed discussion of electrical lines will be reserved for Chap. IX. However, we shall examine the problem in sufficient detail here to form a basis for an analogy with mechanical lattices.

In the last section we found that the quantity $\sqrt{1/LC}$ played the part for electrical lines that $\sqrt{U''/M}$ plays for mechanical lattices. The classical method for drawing an analogy between electromagnetic and mechanical effects is to associate electromagnetic energy with kinetic energy and electrostatic energy with potential energy. This leads to associating

$$\frac{1}{C} \text{ with } U'' \quad \text{and} \quad L \text{ with } M \quad (12.1)$$

However, this method is not the only one that can be used, and we shall find another method more convenient for some purposes. The design of the system under consideration will, in general, determine the analogy to be used.

Another way in which we could make the analogy would be to take

$$U''_p \sim \frac{1}{L_p} \quad \text{and} \quad M \sim C \quad (12.2)$$

For instance, this is the proper analogy to use if we wish to construct an electrical line with the same propagation properties as a lattice with equally spaced particles of equal mass and interactions between all particles. This can best be shown by con-

structing such a line according to Eq. (12.2) and verifying that the line equations of the two systems are exactly the same. The line is shown in Fig. 12.1. Each condenser has capacity C and is connected to its nearest neighbors through an inductance L_1 . The condensers are connected to next nearest neighbors by inductances L_2 and to the p th neighbors by inductances L_p . Only L_1 and L_2 are shown in the diagram in order not to complicate it too much. The condensers are numbered as before. The current flowing through L_1 will be $i_{n-1,n}$, $i_{n,n+1}$, $i_{n+1,n+2}$, and, in general, that flowing through L_p will be $i_{n-p,n}$, $i_{n-p+1,n+1}$, \dots ,

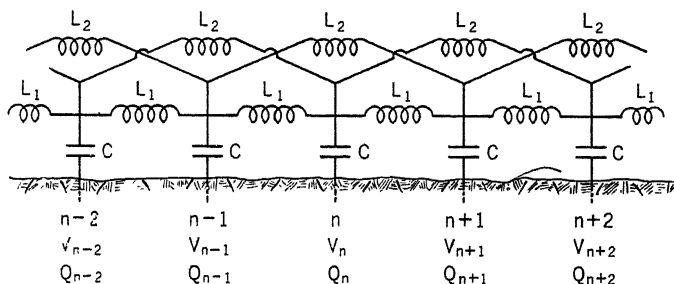


FIG. 12.1.

$i_{n-1,n+p-1}$, $i_{n,n+p}$. The second subscript on the current indicates the condenser into which the current flows, and the first subscript indicates the condenser from which the current started. The charge Q_n on condenser n will be given by

$$\begin{aligned} \frac{dQ_n}{dt} &= i_{n-p,n} + i_{n-p+1,n} + \dots + i_{n-1,n} \\ &\quad - (i_{n,n+1} + i_{n,n+2} + \dots + i_{n,n+p}) \\ &= \sum_p (i_{n-p,n} - i_{n,n+p}) \end{aligned} \quad (12.3)$$

We have the following equations for the current in the various branches of the circuit denoted by L_p , if we take the potential of condenser n to be V_n :

$$\left. \begin{aligned} L_1 \frac{d}{dt} i_{n-1,n} &= V_{n-1} - V_n = \frac{Q_{n-1} - Q_n}{C} \\ L_2 \frac{d}{dt} i_{n-2,n} &= V_{n-2} - V_n = \frac{Q_{n-2} - Q_n}{C} \\ L_p \frac{d}{dt} i_{n-p,n} &= V_{n-p} - V_n = \frac{Q_{n-p} - Q_n}{C} \end{aligned} \right\} \quad (12.4)$$

Differentiating Eq. (12.3) and combining with Eq. (12.4), we obtain

$$C \frac{d^2 Q_n}{dt^2} = C \sum_p \left(\frac{di_{n-p,n}}{dt} - \frac{di_{n,n+p}}{dt} \right) = \sum_p \frac{Q_{n-p} + Q_{n+p} - 2Q_n}{L_p} \quad (12.5)$$

Equation (12.5) is indeed identical with that for a row of particles, each having mass M , with interactions allowed among all neighbors [Eq. (8.8)], if we make the correlation.

$$M \sim C \quad \text{and} \quad U''_p \sim \frac{1}{L_p} \quad (12.2)$$

The line shown in Fig. 12.1 will thus have the same propagation properties as the lattice of like particles with unlimited interactions (Chap. III, Sec. 8).

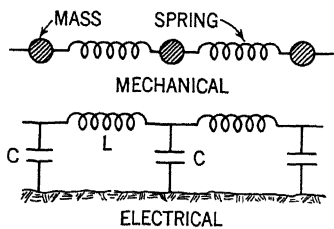


FIG. 12.2.

A geometrical argument leading to Eq. (12.2) may be given. The mechanical low-pass filter consists of point masses joined by elastic elements that we might visualize as springs. The elastic elements (Fig. 12.2) each have two ends, one connected to one

mass and one to another mass, while the masses are represented by single points. An electric line having all its condensers shunting the high frequencies may be regarded as a single line with the condensers connected between the line and ground at regular intervals. Then the inductances appear as having two ends connected to different condensers, and the condensers are essentially points in the structure. Another way of looking at the problem is to regard the elastic forces as coupling forces in the lattice and the inductances as coupling forces in the electric line, while the masses and condensers are thought of as supplying inertial forces to their respective systems.

In the case of a high-pass filter, the electric circuit would have inductances leading to ground with condensers incorporated in the line and separating the inductances. In this case the inductances would have to be regarded as the points of the system and

the condensers as the parts having two ends, so that the classical analogy [Eq. (12.1)] would again hold. For a band-pass filter with a low-pass band and higher bands in addition, the inductances would have to be shunted by condensers that would be regarded as masses, since one plate of each condenser could still be taken as grounded. However, a closer analysis of the system would be necessary to decide which analogy to use, since there might be condensers elsewhere in the circuit.

There is a limit to which these analogies may be carried. It is not possible, for instance, to construct an electrical line by Eq. (12.5), giving an arbitrary relation between α and ν , as it is for a mechanical lattice (discussed in Sec. 10). The reason is that it is sometimes necessary to allow U''_p to take on negative values. This is easy to realize mechanically, but it would not be possible to obtain a negative self-inductance for the analogous electrical line.

CHAPTER IV

MATHEMATICAL TREATMENT OF MORE COMPLICATED ONE-DIMENSIONAL LATTICES

13. Equations of Motion for the One-dimensional NaCl Lattice

The one-dimensional NaCl lattice is a special case of the one-dimensional diatomic lattice that was discussed qualitatively in Secs. 3 and 7. The general lattice is shown in Fig. 13.1. There are two masses M_1 and M_2 alternating. A given mass M_1 will have its right-hand neighbor a distance d_1 away and its left-hand

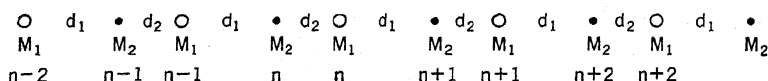


FIG. 13.1.—A row of diatomic molecules.

neighbor a distance d_2 on the other side. The period of the lattice is then

$$d = d_1 + d_2 \quad (13.1)$$

In Sec. 7 we assumed one mass, say M_2 , much smaller than the other. Then M_1 was supposed to interact with the small mass nearest to it and with each of the two large masses nearest to it. The small masses were supposed to interact only with the nearest large mass. In other words, we allowed molecules as a whole to interact and then included the internal degree of freedom in our discussion.

In this section we shall discuss a slightly different lattice. The two will have the same type of curve, however, since we shall change only the rules of interaction. The *interactions* shall take place *between nearest neighbors only*, without reference to the size of the masses. This implies, of course, that we are dealing with particles that are comparable. If we limit the problem to one in which the distances are equal and the interactions of a particle with its two nearest neighbors are equal, we obtain the one-dimensional analogue of the NaCl lattice used by Born in his

theory of specific heats. The lattice is shown in Fig. 13.2. The solid dots represent particles of mass M_2 and the open circles those of mass M_1 . The particles can be numbered in two different ways as shown in Figs. 13.1 and 13.2. We use the second one, where we have assigned even numbers to solid dots and odd numbers to the open circles. This means that the equilibrium coordinates of the particles with mass M_1 are $(2n + 1)d/2$, while the equilibrium coordinates of particles with mass M_2 are $2nd/2 = nd$.

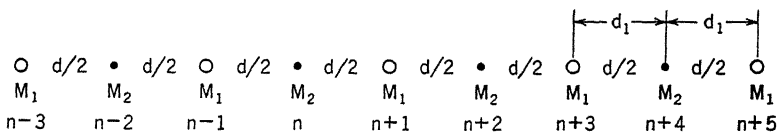


FIG. 13.2.—M. Born's model for sodium chloride.

The equations of motion of the two types of particles are different because of their different masses. If we denote the force on the m th particle by F_m , which is computed exactly as in Sec. 8, Eq. (8.6) or Eq. (11.4), we obtain for the equations of motion

$$\left. \begin{aligned} F_{2n} &= U''_1(y_{2n-1} + y_{2n+1} - 2y_{2n}) = M_2 \frac{d^2 y_{2n}}{dt^2} \\ F_{2n+1} &= U''_1(y_{2n} + y_{2n+2} - 2y_{2n+1}) = M_1 \frac{d^2 y_{2n+1}}{dt^2} \end{aligned} \right\} \quad (13.2)$$

where y_k is the displacement of the k th particle from its equilibrium position. Let us assume a wave solution to these equations of the following form:

$$\left. \begin{aligned} y_{2n} &= A_2 e^{i(\omega t - 2nk_1)} \\ y_{2n+1} &= A_1 e^{i(\omega t - (2n+1)k_1)} \end{aligned} \right\} \quad (13.3)$$

where

$$\begin{aligned} k &= 2\pi a d \\ k_1 &= 2\pi a \frac{d}{2} = \pi a d = \frac{1}{2} k \\ \omega &= 2\pi \nu \\ a &= \frac{1}{\lambda} \end{aligned}$$

It should be noted that the first of Eqs. (13.3) represents a wave propagating only through the particles of mass M_2 , while the second represents a wave propagating only through those of

mass M_1 . The wave lengths and frequencies for a given disturbance must be equal. The amplitudes of the two waves, on the other hand, are not necessarily equal. They may differ in magnitude as well as in phase.

In order that Eq. (13.3) may satisfy Eq. (13.2), certain relations must be imposed on the constants in the solution. These relations are obtained by substituting the assumed solution (13.3) in Eq. (13.2). The substitution yields

$$\begin{aligned} M_2(-A_2\omega^2) &= U''_1(A_1e^{ik_1} + A_1e^{-ik_1} - 2A_2) \\ M_1(-A_1\omega^2) &= U''_1(A_2e^{ik_1} + A_2e^{-ik_1} - 2A_1) \end{aligned}$$

The exponential term $e^{i(\omega t - 2nk_1)}$ divides out of the first equation, while $e^{i(\omega t - (2n+1)k_1)}$ divides out of the second. Making use of the relation

$$e^{ik_1} + e^{-ik_1} = 2 \cos k_1$$

and rearranging terms, we obtain two linear equations in A_1 and A_2 .

$$\left. \begin{aligned} A_2(M_2\omega^2 - 2U''_1) + 2A_1U''_1 \cos k_1 &= 0 \\ A_1(M_1\omega^2 - 2U''_1) + 2A_2U''_1 \cos k_1 &= 0 \end{aligned} \right\} \quad (13.4)$$

The condition that these equations give nontrivial solutions for A_1 and A_2 is that the determinant of the coefficients of A_1 and A_2 shall vanish. This condition gives us a relation between ω and k_1 in terms of the constants of the lattice: M_1 , M_2 , and U''_1 . Thus

$$(M_1\omega^2 - 2U''_1)(M_2\omega^2 - 2U''_1) = 4U''_1{}^2 \cos^2 k_1$$

or, expanding,

$$\omega^4 - 2U''_1 \left(\frac{1}{M_1} + \frac{1}{M_2} \right) \omega^2 + 4 \frac{U''_1{}^2}{M_1 M_2} \sin^2 k_1 = 0 \quad (13.5)$$

This equation possesses two solutions for ω^2 and hence two solutions for ω , since the frequency is always taken to be positive; *i.e.*, for each value of k_1 there will be two values of the frequency, so that the ω vs. k_1 curve will have two branches.

$$\omega^2 = U''_1 \left[\left(\frac{1}{M_1} + \frac{1}{M_2} \right) \pm \sqrt{\left(\frac{1}{M_1} + \frac{1}{M_2} \right)^2 - 4 \frac{\sin^2 k_1}{M_1 M_2}} \right] \quad (13.6)$$

Substitution of Eq. (13.6) into Eq. (13.4) yields two equations for A_1 and A_2 . These two equations are, however, not linearly

independent and hence may be used only to determine the ratio A_1/A_2 , which is real. The magnitudes and actual phases of the amplitudes for the two waves will depend on the initial conditions.

14. Electrical Analogue of the One-dimensional Diatomic Lattice

To construct the electrical line analogous to the one-dimensional diatomic lattice, we must use the classical method of association [Eq. (12.1)]. This means that since we have two masses in the mechanical model, we must have two inductances in the electrical model. We could generalize the problem treated in the preceding section and allow different coupling between the two masses or, what amounts to the same thing, allow the distance between M_1 and M_2 to be different on the two sides of the particle.

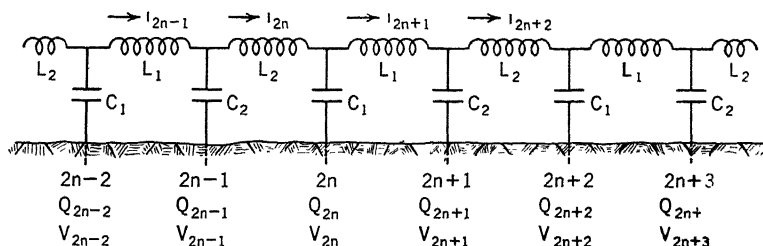


FIG. 14.1.—Electric line corresponding to the sodium-chloride model.

This would give an electric line with condensers C_1 and C_2 alternating. The condenser C_1 to the right of a given condenser C_2 would be joined to it by an inductance L_2 , while the condenser C_1 to the left would be joined by an inductance L_1 . This arrangement would, in general, be analogous to the mechanical model described in Sec. 7.

The electric line is shown in Fig. 14.1. As before, i_m represents current flowing from condenser $(m - 1)$ to condenser m as in the case of Fig. 11.3. The fundamental equations are

$$i_{2n} - i_{2n+1} = \frac{dQ_{2n}}{dt}; \quad i_{2n+1} - i_{2n+2} = \frac{dQ_{2n+1}}{dt} \quad (14.1)$$

$$\left. \begin{aligned} L_1 \frac{di_{2n+1}}{dt} &= V_{2n} - V_{2n+1} = \frac{Q_{2n}}{C_1} - \frac{Q_{2n+1}}{C_2} \\ L_2 \frac{di_{2n}}{dt} &= V_{2n-1} - V_{2n} = \frac{Q_{2n-1}}{C_2} - \frac{Q_{2n}}{C_1} \end{aligned} \right\} \quad (14.2)$$

Differentiating Eq. (14.2) and combining with Eq. (14.1) will yield

$$\left. \begin{aligned} L_1 \frac{d^2 i_{2n+1}}{dt^2} &= \frac{i_{2n} - i_{2n+1}}{C_1} - \frac{i_{2n+1} - i_{2n+2}}{C_2} \\ L_2 \frac{d^2 i_{2n}}{dt^2} &= \frac{i_{2n-1} - i_{2n}}{C_2} - \frac{i_{2n} - i_{2n+1}}{C_1} \end{aligned} \right\} \quad (14.3)$$

These two equations would be identical with Eq. (13.2) for the diatomic lattice treated in the last section if $C_1 = C_2$ and we replaced capacitance by the elastic constant and inductance by mass.

The solution of Eq. (14.3) is carried out in exactly the same way as that of Eq. (13.2). We assume wave solutions for i_{2n} and i_{2n+1} with the same frequency and wave number but with different amplitudes, as in Eq. (13.3).

$$i_{2n} = A_2 e^{i(\omega t - 2n k_1)}, \quad i_{2n+1} = A_1 e^{i[\omega t - (2n+1)k_1]} \quad (14.4)$$

Substitution in Eq. (14.3) gives two equations linear in the amplitudes

$$\left. \begin{aligned} \left(-L_1 \omega^2 + \frac{1}{C_1} + \frac{1}{C_2}\right) A_1 - \left(\frac{1}{C_1} e^{ik_1} + \frac{1}{C_2} e^{-ik_1}\right) A_2 &= 0 \\ \left(-L_2 \omega^2 + \frac{1}{C_1} + \frac{1}{C_2}\right) A_2 - \left(\frac{1}{C_2} e^{ik_1} + \frac{1}{C_1} e^{-ik_1}\right) A_1 &= 0 \end{aligned} \right\} \quad (14.5)$$

These simultaneous linear equations in A_1 and A_2 have a non-trivial solution if their determinant vanishes.

$$\begin{aligned} &\left(-L_1 \omega^2 + \frac{1}{C_1} + \frac{1}{C_2}\right) \left(-L_2 \omega^2 + \frac{1}{C_1} + \frac{1}{C_2}\right) \\ &\quad - \left(\frac{1}{C_1} e^{ik_1} + \frac{1}{C_2} e^{-ik_1}\right) \left(\frac{1}{C_2} e^{ik_1} + \frac{1}{C_1} e^{-ik_1}\right) = 0 \end{aligned} \quad (14.6)$$

which reduces to

$$\omega^4 - \omega^2 \left(\frac{1}{L_1} + \frac{1}{L_2}\right) \left(\frac{1}{C_1} + \frac{1}{C_2}\right) + \frac{\sin^2 k_1}{L_1 L_2 C_1 C_2} = 0 \quad (14.7)$$

the solution of which is

$$\begin{aligned} \omega^2 &= \frac{1}{2} \left(\frac{1}{L_1} + \frac{1}{L_2}\right) \left(\frac{1}{C_1} + \frac{1}{C_2}\right) \\ &\quad \pm \sqrt{\frac{1}{4} \left(\frac{1}{L_1} + \frac{1}{L_2}\right)^2 \left(\frac{1}{C_1} + \frac{1}{C_2}\right)^2 - \frac{4 \sin^2 k_1}{L_1 L_2 C_1 C_2}} \end{aligned} \quad (14.8)$$

This reduces to the expression (13.6) obtained for the mechanical case if $L_1 = M_2$, $L_2 = M_1$, and $1/U''_1 = C_1 = C_2$. There will be two branches to the ω vs. k_1 curve whether $L_1 \neq L_2$ or not, but taking $C_1 \neq C_2$ would distort the shape of the curves.

This problem was discussed by electrical engineers¹ who did not notice the similarity with the one-dimensional NaCl lattice discussed by Born. The problem originated from an attempt to join an aerial telephonic line with a city cable, as shown in

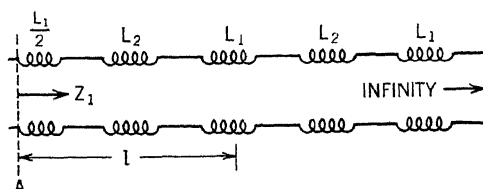
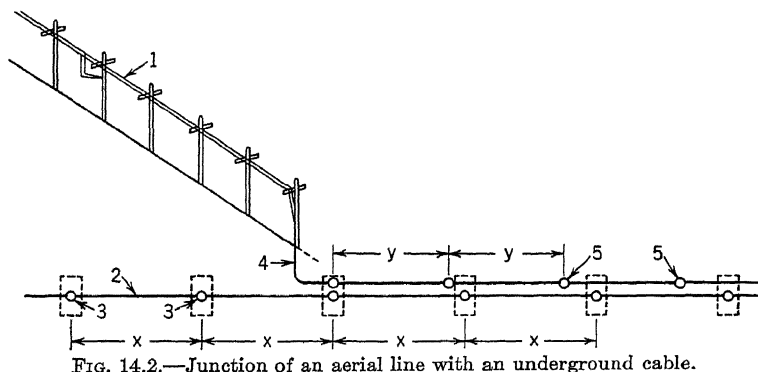


Fig. 14.2. In order to obtain a correct junction at 4, where the line is connected with the cable, it would be necessary to load the cable with equal coils at a distance y, y, \dots . This results from two conditions that must be satisfied in order to match the line and the cable at their junction: (1) to have the same passing bands, and (2) to have the same characteristic impedances (see Chap. V). The difficulty was that the underground city cable was already built to receive its loading coils at given distances x, x, \dots . The solution proposed consists in using alternately two types of coils L_1 and L_2 (Fig. 14.3), resulting in a structure

¹ FRENCH, N. R., U.S. patent 1,741,926, Dec. 31, 1929; S. P. MEAD and N. R. FRENCH, U.S. patent 1,769,959, July 8, 1930.

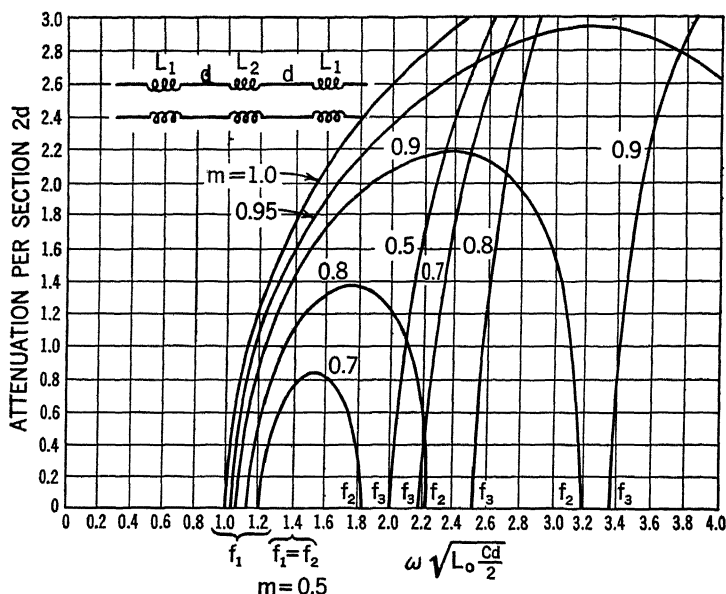


FIG. 14.4.—Curves computed by Mead and French. Compare with Fig. 7.3 or 3.9b.

practically identical with the one of Fig. 14.1. Attenuation curves for different values of m [$L_1 = mL_0$, $L_2 = (1 - m)L_0$, L_0 a constant] were computed and are shown in Fig. 14.4. They are identical with the attenuation curves β shown in Fig. 3.9b, which were obtained by Born for the NaCl structure, the theory of which will now be discussed.

15. Discussion of the One-dimensional NaCl Lattice

In this section we shall discuss the motion given by the two branches of the ω vs. $k_1 = k/2$ curve with particular attention to the case $k_1 = 0$ and $k_1 = \pm\pi/2$. The relation between ω and k_1 is given by Eq. (13.6).

$$\omega^2 = U''_1 \left[\left(\frac{1}{M_1} + \frac{1}{M_2} \right) \pm \sqrt{\left(\frac{1}{M_1} + \frac{1}{M_2} \right)^2 - \frac{4 \sin^2 k_1}{M_1 M_2}} \right] \quad (13.6)$$

or, rearranging terms,

$$\omega^2 = \frac{U''_1}{M_1 M_2} (M_1 + M_2 \pm \sqrt{M_1^2 + M_2^2 + 2M_1 M_2 \cos 2k_1}) \quad (15.1)$$

Equation (15.1) is completely symmetrical in M_1 and M_2 , and we may therefore assume M_1 the larger of the two masses without loss of generality.

$$M_1 > M_2$$

The ratio of the amplitudes of the waves may be obtained from either of Eqs. (13.4). Both give the same result in terms of k_1 . Using the first,

$$\frac{A_1}{A_2} = \frac{2U''_1 - M_2\omega^2}{2U''_1 \cos k_1} \quad (15.2)$$

and, substituting Eq. (15.1) for ω^2 , we obtain

$$\frac{A_1}{A_2} = \frac{M_1 - M_2 \mp \sqrt{M_1^2 + M_2^2 + 2M_1M_2 \cos 2k_1}}{2M_1 \cos k_1} \quad (15.3)$$

The minus sign in Eq. (15.3) corresponds to the plus sign in Eq. (15.1) or the upper branch of the ω vs. k_1 curve, while the plus sign of Eq. (15.3) corresponds to the minus sign of Eq. (15.1) or the lower branch. It should be noted that the amplitude ratio is always real; therefore, the waves may have only two phase relations: phase difference zero if $A_1/A_2 > 0$, and phase difference π if $A_1/A_2 < 0$. This is typical of a system without any resistance and with no damping.

For large wave lengths λ , $k_1 \rightarrow 0$ as does k . For this case we may set

$$\cos 2k_1 \approx 1 - 2k_1^2 = 1 - \frac{1}{2}k^2 \approx \cos k$$

where

$$k_1 = 2\pi ad_1 = \pi ad = \frac{1}{2}k$$

and the radical in Eq. (15.1) becomes

$$\begin{aligned} & \sqrt{M_1^2 + M_2^2 + 2M_1M_2 \left(1 - \frac{k^2}{2}\right)} \\ &= (M_1 + M_2) \sqrt{1 - \frac{k^2 M_1 M_2}{(M_1 + M_2)^2}} \\ &\approx (M_1 + M_2) \left(1 - \frac{1}{2} \frac{k^2 M_1 M_2}{(M_1 + M_2)^2}\right) \end{aligned} \quad (15.4)$$

and Eq. (15.1) reduces to

$$\left. \begin{aligned} \omega_-^2 &= \frac{k^2 U''_1}{2(M_1 + M_2)} && \text{lower branch} \\ \omega_+^2 &= 2U''_1 \left[\frac{1}{M_1} + \frac{1}{M_2} - \frac{k^2}{4(M_1 + M_2)} \right] && \text{upper branch} \end{aligned} \right\} \quad (15.5)$$

The subscripts $+$ and $-$ denote the sign used before the radical. Thus ω_- is linear in k near the origin as in the case of like particles; ω_+ has a maximum at the origin and decreases parabolically as $|k|$ increases.

To interpret properly the meaning of a second frequency for infinite wave length, we must compute the amplitude ratio for small k_1 . Substitution of Eq. (15.4) into Eq. (15.3) yields the following relations for small k (powers of k higher than the second are neglected):

$$\left. \begin{aligned} \left(\frac{A_1}{A_2}\right)_+ &= 1 + \frac{k^2}{8} \frac{M_1 - M_2}{M_1 + M_2} && \text{lower branch} \\ \left(\frac{A_1}{A_2}\right)_- &= -\frac{M_2}{M_1} \left(1 - \frac{k^2}{8} \frac{M_1 - M_2}{M_1 + M_2}\right) && \text{upper branch} \end{aligned} \right\} \quad (15.6)$$

Thus the lower branch increases parabolically at the origin as $|k|$ increases from zero. At $k = 0$

$$\left(\frac{A_1}{A_2}\right)_+ = 1, \quad \left(\frac{A_1}{A_2}\right)_- = -\frac{M_2}{M_1}; \quad |k| \ll 1 \quad (15.7)$$

The waves corresponding to the lower branch have equal amplitudes and phase difference zero; thus all the particles are displaced by the same amount and in the same direction. The wave length of each of the waves is infinite, and the lattice is displaced as a whole. There is thus no restoring force, and the frequency is zero. On the other hand, the waves for the upper branch are exactly out of phase; *i.e.*, the displacement of particles of mass M_1 is opposite to that of the neighboring particles M_2 . Evidently the center of mass of two neighboring particles is stationary, but restoring forces enter in so that the frequencies of the waves are no longer zero. The lengths of the waves are still infinite since each wave is regarded as propagating through just one type of particle.

The values for k on the limits of the interval to which k is restricted are $\pm\pi$. The two limits will be symmetrical, and we consider only the case

$$k = \pi - \epsilon = 2k_1 \quad \epsilon \text{ small}$$

Then

$$\cos k = \cos (\pi - \epsilon) = -\cos \epsilon \approx -1 + \frac{\epsilon^2}{2}$$

and the radical in Eq. (15.1) becomes

$$\sqrt{\dots} = (M_1 - M_2) \left[1 + \frac{\epsilon^2 M_1 M_2}{2(M_1 - M_2)^2} \right]$$

if $(M_1 - M_2)$ is not too small. Substitution in Eq. (15.1) yields

$$\left. \begin{aligned} \omega_+^2 &= \frac{2U''_1}{M_2} + \frac{U'''_1 \epsilon^2}{2(M_1 - M_2)} \\ \omega_-^2 &= \frac{2U''_1}{M_1} - \frac{U'''_1 \epsilon^2}{2(M_1 - M_2)} \end{aligned} \right\} \quad (15.8)$$

so that the upper branch increases parabolically from

$$\omega = \sqrt{\frac{2U''_1}{M_2}}$$

while the lower branch decreases parabolically from

$$\omega = \sqrt{\frac{2U''_1}{M_1}}$$

as $|\epsilon|$ increases from zero. It should be noted that $\omega_+ > \omega_-$ at the limits of the interval since $M_1 > M_2$, and between these limiting values of ω we have a stopping band to be discussed later.

The amplitude ratio at the ends of the interval is easily obtained. We have

$$\cos k_1 = \cos \left(\frac{\pi}{2} - \frac{\epsilon}{2} \right) \approx \frac{\epsilon}{2}$$

(since $k = 2k_1$), and therefore from Eq. (15.3)

$$\begin{aligned} \left(\frac{A_1}{A_2} \right)_- &= \frac{M_1 - M_2 - (M_1 - M_2) \left[1 + \epsilon^2 \frac{M_1 M_2}{2(M_1 - M_2)^2} \right]}{\epsilon M_1} \\ &= \frac{-\epsilon M_1 M_2}{2M_1(M_1 - M_2)} \rightarrow 0 \end{aligned} \quad (15.9a)$$

$$\begin{aligned} \left(\frac{A_1}{A_2} \right)_+ &= \frac{M_1 - M_2 + (M_1 - M_2) \left[1 + \frac{\epsilon^2 M_1 M_2}{2(M_1 - M_2)^2} \right]}{\epsilon M_1} \\ &= \frac{2(M_1 - M_2) + \frac{\epsilon^2 M_1 M_2}{2(M_1 - M_2)}}{\epsilon M_1} \rightarrow \infty \end{aligned} \quad (15.9b)$$

as $\epsilon \rightarrow 0$. The interpretation of these ratios is not very difficult. We have already seen that for the upper branch $[(A_1/A_2)_-, \omega_+]$

the amplitude ratio is negative and different from zero at and near the origin. Equation (15.9a) shows that it is negative near the ends of the interval $|k| \leq \pi$ and zero at the ends. Then for infinite wave length the particles oscillate in opposite directions, the lighter particles with larger amplitude. As the wave length decreases, the amplitude of the heavy particles decreases, and for the limiting wave length the light particles oscillate while the heavy particles remain at rest.

For the lower branch $[(A_1/A_2)_{-}, \omega_{-}]$, on the other hand, the particles start out all in phase and with equal amplitudes for infinite wave length. As the wave length decreases, the amplitude of the light particles decreases, and they remain at rest for

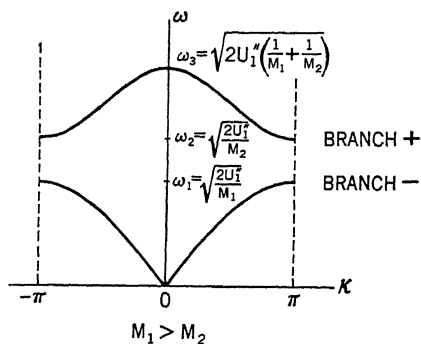


FIG. 15.1.

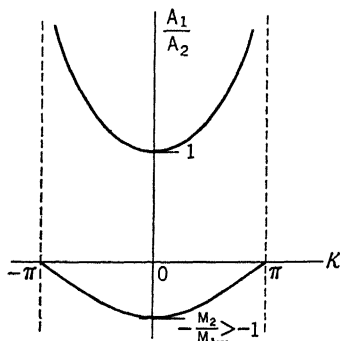


FIG. 15.2.

the limiting wave length while the heavy particles are still oscillating.

These results are summarized in Figs. 15.1 through 15.3. Figure 15.1 shows ω as a function of k for $M_1 > M_2$. Figure 15.2 shows the variation of the amplitude ratio for the two branches, and Fig. 15.3 gives the motion of the particles for the various cases discussed. The arrows in Fig. 15.3 indicate the amplitudes with which the two types of particles oscillate. Figure 15.3 shows clearly that the motions obtained for ν_1 and ν_2 are very similar: for ν_1 the particles M_2 are all at rest, and particles M_1 move in alternate directions. For ν_2 , M_1 is at rest and M_2 moving. The forces involved are the same in both cases, since changes in the distances between particles are the same; hence, the frequency ratio must be proportional to the square root of the inverse ratio of the masses.

$$\frac{\nu_1}{\nu_2} = \sqrt{\frac{M_2}{M_1}}$$

as is actually obtained.

The lower branch is frequently called the *acoustical* branch. This name comes from the fact that the frequencies in it are of the same order of magnitude as acoustical or supersonic vibrations. The upper branch is frequently called the *optical* branch, because of the fact that its frequencies are of the order of magnitude of infrared frequencies. Further, if we think of the lattice

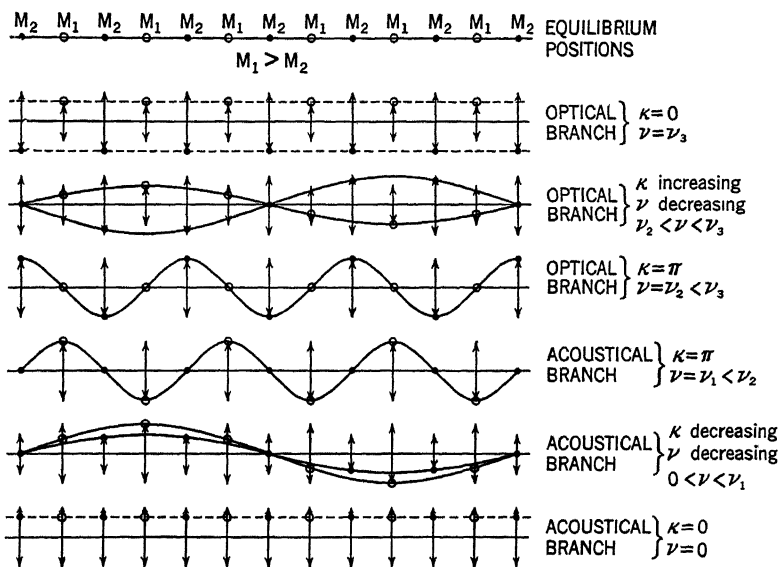


FIG. 15.3.

as being composed of ions having alternate signs, *e.g.*, Na^+ ions alternating with Cl^- ions, an alternating electric field could not excite the acoustical type of wave in which two neighboring particles are in phase, but it could excite the optical type and displace neighboring particles in opposite directions.

So far we have discussed only the passing bands of our lattice. We now consider the stopping bands. These occur for frequencies between ω_1 and ω_2 and for frequencies above ω_3 . We return to Eq. (13.5).

$$-\left(\frac{\omega^2}{2U''_1}\right)^2 M_1 M_2 + \frac{\omega^2}{2U''_1} (M_1 + M_2) = \sin^2 k_1 = \sin^2 \frac{k}{2}$$

We may rewrite this in the form

$$\omega \left(\frac{M_1 + M_2}{2\tilde{t}''_1} - \frac{\omega^2 M_1 M_2}{4\tilde{t}''_1^2} \right) = \sin^2 \frac{k}{2} \quad (15.10)$$

We have seen that as ω increases from zero to $\omega_1 = 2\pi\nu_1$, the expression on the right increases from zero to one. If ω increases still further, the expression on the right becomes greater than one, and $k/2$ must become complex. Let

$$k = \alpha + i\beta \quad (15.11)$$

Then

$$\sin \frac{k}{2} = \sin \frac{\alpha}{2} \cosh \frac{\beta}{2} + i \cos \frac{\alpha}{2} \sinh \frac{\beta}{2} \quad (15.12)$$

and since this expression must be real, we have the condition that

$$\cos \frac{\alpha}{2} \sinh \frac{\beta}{2} = 0 \quad \text{or} \quad \frac{\alpha}{2} = \frac{\pi}{2} \quad (15.13)$$

That is, R.P. $k = \pi = 2\pi ad$ so that R.P. $a = 1/2d$ throughout the stopping band $\omega_1 < \omega < \omega_2$. R.P. means "the real part of." Somewhere between ω_1 and ω_2 the expression on the right of Eq. (15.10) reaches a maximum and starts to decrease. It equals one at ω_2 and is positive and less than one between ω_2 and ω_3 . At ω_3 it is zero, and as ω increases still further, it becomes negative. In other words, Eq. (15.12) becomes pure imaginary and therefore

$$\sin \frac{\alpha}{2} \cosh \frac{\beta}{2} = 0 \quad \text{or} \quad \frac{\alpha}{2} = 0 \quad (15.14)$$

This means that

$$k = i\beta = 2\pi ad$$

Hence

$$\text{R.P. } a = 0 \quad \omega > \omega_3$$

Since the real part of k is constant throughout both stopping bands and only the imaginary part varies, we have attenuation of the waves. In the first case $\omega_1 < \omega < \omega_2$ at the low frequency end ω_1 of the stopping band, the light particles are at rest and the heavy particles are in motion, neighboring heavy particles being just out of phase; and at the other end ω_2 the heavy particles are at rest with the light particles vibrating out of phase. The motion is attenuated along the lattice (*i.e.*, the amplitude of the vibrations decreases from particle to particle) with an

attenuation constant that first increases with the frequency. Somewhere in the stopping band the motion changes from acoustical type to optical type, and as ω increases the attenuation decreases until $\omega = \omega_2$, where it becomes zero.

In the other stopping band $\omega > \omega_3$ the particles are vibrating in opposite phase with the limiting wave length. This motion is attenuated with an attenuation coefficient that increases as ω increases.

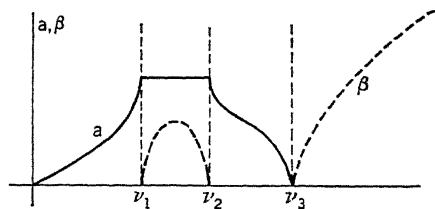


FIG. 15.4.

Curves of wave number a and of attenuation coefficient β against frequency are shown in Fig. 15.4. (Compare with Fig. 14.4.)

16. Transition from a Diatomic to a Monatomic Lattice

The diatomic lattice discussed in the last section is exactly like the monatomic lattice discussed previously except that two masses appear instead of only one; *i.e.*, the distances between neighboring particles are all the same and the interactions are restricted to nearest neighbors. The diatomic lattice may be reduced to a monatomic lattice in three ways:

1. Let $M_2 \rightarrow 0$.
2. Let $M_1 \rightarrow \infty$.
3. Let $M_1 \rightarrow M_2$.

The first two methods leave the period d of the lattice unchanged, while the last halves the period and results in a lattice $d/2 = d_1$. We shall discuss the three methods in the order given above.

1. Let $M_2 \rightarrow 0$.—In this case $\omega_1 = \sqrt{2U''_1/M_1}$ is unchanged, while $\omega_2 = \sqrt{2U''_1/M_2}$ and $\omega_3 = \sqrt{\omega_1^2 + \omega_2^2}$ both go to infinity. The width of the upper passing band goes to zero; for

$$\begin{aligned} \omega_3 - \omega_2 &= \sqrt{\omega_1^2 + \omega_2^2} - \omega_2 = \omega_2 \left(\sqrt{\frac{\omega_1^2}{\omega_2^2} + 1} - 1 \right) \\ &\rightarrow \omega_2 \frac{1}{2} \left(\frac{\omega_1}{\omega_2} \right)^2 \rightarrow 0 \quad (16.1) \end{aligned}$$

Thus the upper band rises and becomes narrower, finally disappearing entirely. The lower branch remains, and we have a low-pass filter with period $d = 2d_1$ left.

2. Let $M_1 \rightarrow \infty$.—Here $\omega_1 = \sqrt{2U''_1/M_1}$ goes to zero. ω_2 remains unchanged and $\omega_3 \rightarrow \omega_2$. Thus in the limiting case there is only a single frequency $\omega_2 = \sqrt{2U''_1/M_2}$, and this frequency does not really propagate. Each of the light particles oscillates separately with frequency ω_2 . This corresponds to the case of a row of harmonic oscillators with no interaction. The heavy masses are responsible for the restoring force on the oscillators but take no part in the motion themselves. The amplitude

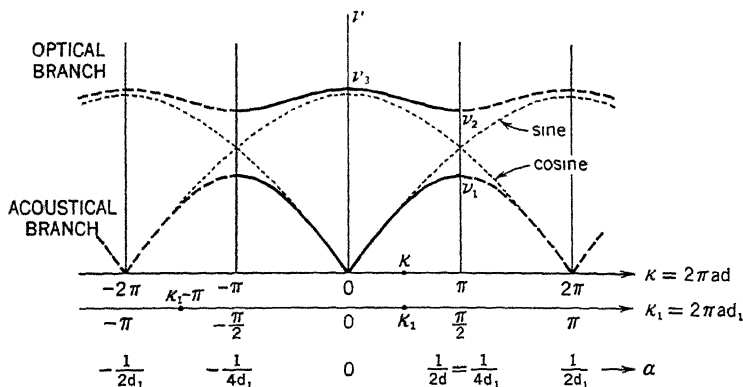


FIG. 16.1.

of the vibration is, of course, restricted to values less than d_1 ; the light particles must not go through the heavy particles.

Had we allowed interactions between second neighbors as well as nearest neighbors, we would have obtained in the limiting case a lattice of coupled harmonic oscillators that would lead to a band-pass filter. The single frequency present for independent oscillators would spread out into a band; the lower branch present in the diatomic lattice would still be missing.

3. Let $M_1 \rightarrow M_2$.—This process is considerably more complicated than the previous two because a sudden change in the periodicity of the lattice is involved. The original structure, with $M_1 > M_2$, repeats itself after a distance d , but when $M_1 = M_2$, the period suddenly drops to $d_1 = d/2$. Let us first discuss the relation between frequency and wave number $\alpha = 1/\lambda$. This relation was shown in Fig. 15.1, which must be

understood as representing only one section of a periodic curve, as drawn in Fig. 16.1. The central section (Fig. 15.1) corresponds to $-\pi < k < \pi$ where $k = 2\pi ad$ as usual, and the complete curve is obtained when k takes any arbitrary value.

When $M_1 = M_2$, two changes must be made:

a. The change in periodicity results in a sudden extension of the fundamental interval. For a lattice with period d , the wave number a has period $1/d$, and its fundamental interval extends from $-1/2d$ to $+1/2d$. When the lattice period changes to $d_1 = d/2$, the wave-number period becomes $1/d_1 = 2/d$, and the fundamental interval is $\pm 1/2d_1 = \pm 1/d$.

The following table summarizes the changes in a , k , and k_1 :

	Period for				Fundamental interval	
	Lattice	a	k	k_1	k	k_1
$M_1 > M_2$	d	$\frac{1}{d}$	2π	π	$\pm\pi$	$\pm\frac{\pi}{2}$
$M_1 = M_2$	$d_1 = \frac{d}{2}$	$\frac{1}{d_1} = \frac{2}{d}$	4π	2π	$\pm 2\pi$	$\pm\pi$

(16.2)

where $k = 2\pi ad$ and $k_1 = 2\pi ad_1 = k/2$.

b. Another change in the curve is that it must become a single curve as in Fig. 2.4 instead of the double curve of Fig. 15.1. The single curve is drawn as a dotted line in Fig. 16.1, assuming that $M_1, M_2 \rightarrow M = 2\sqrt{M_1 M_2 / (M_1 + M_2)}$ simultaneously.

All this can be obtained from Eq. (13.6), giving the frequency as a function of k_1 . If we take $M_1 = M_2 = M$, the formula reduces to

$$\frac{M\omega_2}{2U''_1} = 1 \pm \sqrt{1 - \sin^2 k_1} = 1 \pm \cos k_1 = \begin{cases} 2 \sin^2 \frac{k_1}{2} \\ 2 \cos^2 \frac{k_1}{2} \end{cases} \quad (16.3)$$

Selecting the sine function, we obtain

$$\omega = 2 \sqrt{\frac{U''_1}{M}} \left| \sin \frac{k_1}{2} \right| \quad (16.4)$$

which is identical with Eq. (9.1) for the monatomic structure (Fig. 2.4). The cosine curve duplicates the results and in its

middle part represents the upper curve of Fig. 16.1. The sine and cosine curves intersect at a point that is the common limit of ω_1 and ω_2 .

$$\omega_1 \rightarrow \omega_2 \rightarrow \sqrt{\frac{2U''_1}{M}} \quad M_1 \rightarrow M_2 \rightarrow M$$

and the stopping band disappears.

Another aspect of this transformation refers to the description of the wave and of the motion of the particles of the lattice. Referring to Eq. (13.3),

$$\left. \begin{aligned} y_{2n} &= A_2 e^{i(\omega t - 2nk_1)} \\ y_{2n+1} &= A_1 e^{i[\omega t - (2n+1)k_1]} \end{aligned} \right\} \quad (13.3)$$

we see that the solution for the lattice (with $M_1 > M_2$) is represented as two waves, one propagating along particles of mass M_2 and the other propagating along particles of mass M_1 . The wave number k_1 is therefore to be restricted to values between $-\pi/2$ and $+\pi/2$. For the discussion of this section, it will be convenient to *change* our conventions and obtain the solution (13.3) as a *wave propagating through all of the particles*. This means that we must allow k_1 to take on values in the larger interval from $-\pi$ to π . To achieve this we introduce two new quantities C and D , defined by

$$\left. \begin{aligned} A_1 &= C - D = C + D e^{i(2n+1)\pi} \\ A_2 &= C + D = C + D e^{i(2n\pi)} \end{aligned} \right\} \quad (16.5)$$

From Eq. (16.5) it follows that

$$\frac{D}{C} = \frac{A_2 - A_1}{A_2 + A_1}$$

Equation (13.3) may now be written

$$\begin{aligned} y_{2n} &= C e^{i(\omega t - 2nk_1)} + D e^{i[\omega t - 2n(k_1 - \pi)]} \\ y_{2n+1} &= C e^{i[\omega t - (2n+1)k_1]} + D e^{i[\omega t - (2n+1)(k_1 - \pi)]} \end{aligned}$$

and the sum of the two waves

$$y_m = C e^{i(\omega t - mk_1)} + D e^{i[\omega t - m(k_1 - \pi)]} \quad (16.6)$$

gives a single wave propagating through all (both M_1 and M_2) particles. The two methods of representing the wave are shown in Figs. 16.2a to 16.3b. Figure 16.2 shows the representation

with two waves, one passing through each set of particles. The a part is for the acoustical and the b part for the optical branch. Figure 16.3 shows the k_1 and $k_1 - \pi$ waves and their sums for the acoustical and optical branches in a and b , respectively. It

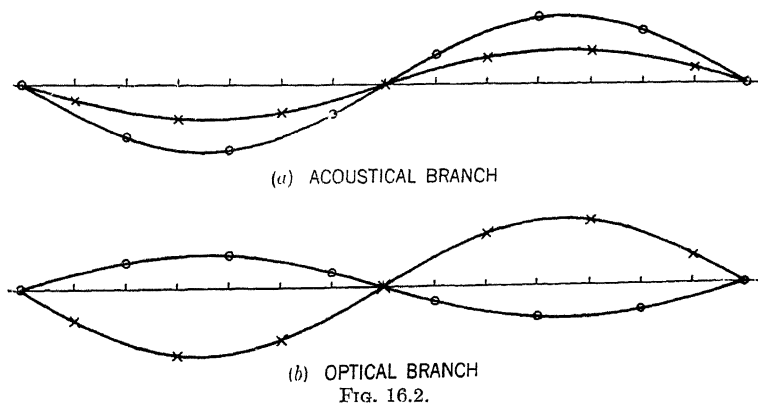


FIG. 16.2.

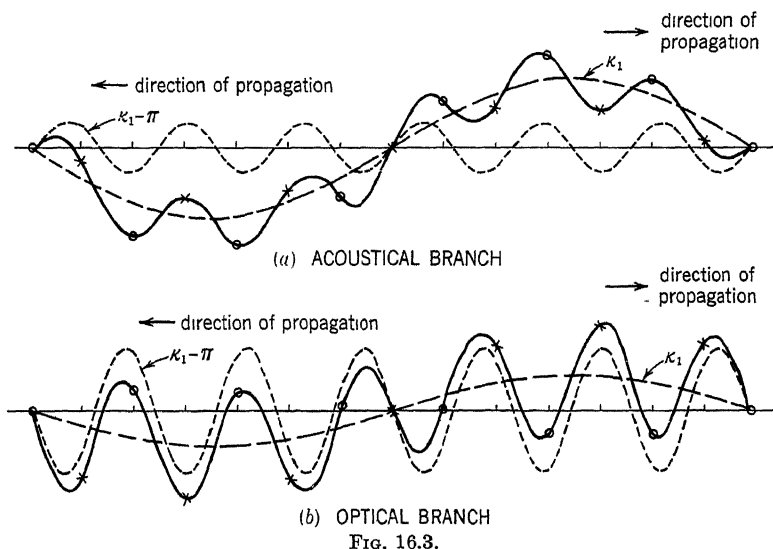


FIG. 16.3.

should be noted that the k_1 and $k_1 - \pi$ waves propagate in opposite directions, so that one may think of the wave propagating to the right as being partially reflected as it traverses each particle, thus giving rise to a disturbance that consists of a transmitted and a reflected wave.

In order to see clearly how the transition from the diatomic to the monatomic lattice takes place, we must refer to Table (16.2), which shows the interval of variation for k_1 in both cases. The original k_1 was restricted to values between $-\pi/2$ and $\pi/2$, which means that $k_1 - \pi$ varies between $-\pi$ and $-\pi/2$ for $k_1 > 0$ or $\pi/2$ and π for $k_1 < 0$, since k_1 and $k_1 + 2\pi$ are equivalent. This

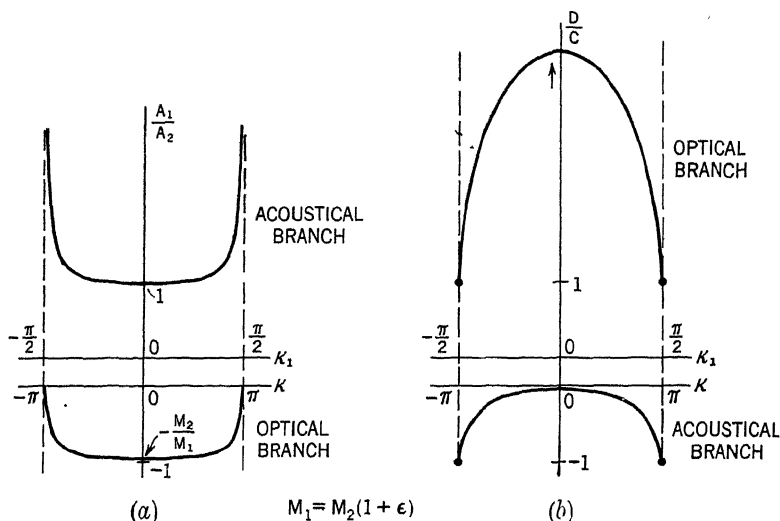


FIG. 16.4.

extends the interval to $-\pi, \pi$ as shown in Table (16.2). The following scheme summarizes this transformation:

$$\begin{array}{c}
 k_1 \quad \begin{array}{c} -\pi/2 \quad \circ \quad \pi/2 \\ \bullet \quad \bullet \quad \bullet \end{array} \\
 k_1 - \pi \quad \begin{array}{c} -\pi \quad -\pi/2 \quad \pi/2 \quad \pi \\ \bullet \quad \bullet \quad \bullet \end{array}
 \end{array}
 \quad (k_1 - \pi + 2\pi = k_1 + \pi) \quad (16.7)$$

This explains the correspondence between the different branches of the curves in Fig. 16.1.

We have previously discussed the variation of the ratio A_1/A_2 for the different types of waves [Eq. (15.3) and Fig. 15.2]. These same curves were drawn again in Fig. 16.4a under the assumption of a very small difference between the masses.

$$M_2 = M_1(1 - \epsilon) \quad \epsilon \ll 1$$

In this case Eq. (15.3) reduces to

$$\begin{aligned} \frac{A_1}{A_2} &= \frac{1 - \frac{M_2}{M_1} \mp \sqrt{1 + \left(\frac{M_2}{M_1}\right)^2 + 2 \frac{M_2}{M_1} \cos 2k_1}}{2 \cos k_1} \\ &= \frac{\epsilon \mp \sqrt{2 + 2(1 - \epsilon) \cos 2k_1 - 2\epsilon + \epsilon^2}}{2 \cos k_1} \\ &= \frac{\frac{\epsilon}{2} \mp \sqrt{\cos^2 k_1 (1 - \epsilon) + \left(\frac{\epsilon}{2}\right)^2}}{\cos k_1} \approx \frac{\epsilon}{2 \cos k_1} \mp \left(1 - \frac{\epsilon}{2}\right) \\ &\quad \cos k_1 \gg \epsilon \quad (16.8) \\ &\quad -\frac{\pi}{2} \leq k_1 \leq \frac{\pi}{2} \end{aligned}$$

since $\frac{1}{2}(1 + \cos 2k_1) = \cos^2 k_1$.

The plus sign gives the acoustical branch, and the minus sign corresponds to the optical branch. The curves remain very near the horizontals ± 1 except at the ends of the interval.

These results can be expressed in terms of the ratio D/C of our new waves [Eq. (16.5)].

Acoustical branch:

$$\frac{D}{C} = \frac{1 - \frac{A_1}{A_2}}{1 + \frac{A_1}{A_2}} = \frac{-\frac{\epsilon}{2 \cos k_1} + \frac{\epsilon}{2}}{2 + \frac{\epsilon}{2 \cos k_1} - \frac{\epsilon}{2}} \approx \frac{\epsilon}{4} \left(1 - \frac{1}{\cos k_1}\right) \quad (16.9)$$

The C wave is dominant with a very small D wave.

Optical branch:

$$\frac{D}{C} = \frac{2 - \frac{\epsilon}{2 \cos k_1} - \frac{\epsilon}{2}}{\frac{\epsilon}{2 \cos k_1} + \frac{\epsilon}{2}} \approx \frac{4}{\epsilon \left(1 + \frac{1}{\cos k_1}\right)} \quad (16.10)$$

The D wave is dominant with a small C wave.

Here we see that in the limit $M_1 = M_2$ the description of the wave motion is much simpler with the C , D waves of Eq. (16.6) than with the A_1 , A_2 waves previously used.

Let us allow k_1 to run from $-\pi$ to $+\pi$ as shown in the diagram 16.7. For the *acoustical branch*

$$-\frac{\pi}{2} < k_1 < \frac{\pi}{2} \quad D \rightarrow 0 \quad C \neq 0 \quad (16.11)$$

while for the *optical branch* we obtain

$$\left. \begin{array}{l} -\pi < k_1 < -\frac{\pi}{2} \\ \frac{\pi}{2} < k_1 < \pi \end{array} \right\} \quad C \rightarrow 0 \quad D \neq 0 \quad (16.12)$$

There is only one wave left (either C or D) almost everywhere except in the immediate neighborhood of $k_1 = \pm\pi/2$, which are the branching points where the curves separate in case $M_1 > M_2$ and give place to a stopping band.

The example just discussed is very important, since it represents the first instance of a general type of problem very often

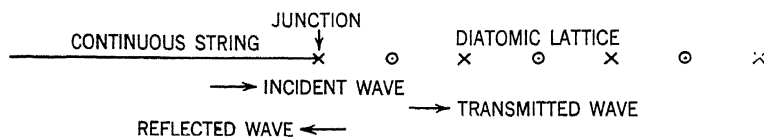


FIG. 16.5.

encountered on other occasions. Here it was possible to follow the transformation from the unperturbed case $M_1 = M_2 = M$ to the perturbed problem $M_1 \neq M_2$ in all details. This is not always possible, and the method followed in more complicated problems will be to start from the unperturbed C, D plane waves and to make linear combinations of them [as in Eq. (16.6)] before discussing the perturbation near the branching points. Such examples may be found in connection with electromagnetic waves (X rays) or with electronic De Broglie waves in crystals, when the periodic distribution of atoms in the crystal lattice can be treated as a small perturbation.

One more remark should be added to show the *connection* between *passing or stopping bands* and *reflection of waves*. If a continuous line capable of transmitting all frequencies is joined to the diatomic lattice (see Fig. 16.5), the coefficient of reflection at the junction will depend on the frequency incident from the continuous line. If the frequency is in one of the stopping bands of the lattice, total reflection will occur; *i.e.*,

$$R = \text{coefficient of reflection} = 1$$

while for a frequency in a passing band both a reflected and a transmitted wave will be excited. The coefficient of reflection will be less than one, and the actual value will depend on the characteristics of the lattice in this case.

17. The One-dimensional Lattice of Polyatomic Molecules

To treat a lattice of polyatomic molecules, we divide the lattice into cells. A cell contains one period of the lattice; *i.e.*, if we start out with atom 1, then the first cell consists of atom 1 to N , where atom $N + 1$ has the same relation to atom $N + 1 + m$ as atom 1 has to atom $1 + m$. Having defined what we mean by *cell* (in general, the same as molecule, unless the molecule itself possesses a periodic structure that is a period of the lattice), we

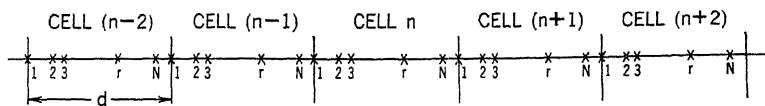


FIG. 17.1.—A row of polyatomic molecules.

change our notation slightly. We number the atoms in a given cell from 1 to N . The cells are also numbered, n being used to denote an arbitrary cell and $n + p$ being the number of the p th cell to the right of cell n . The notation is illustrated in Fig. 17.1. The crosses indicate the equilibrium positions of the atoms, and the vertical lines the positions of the first atom in each cell, *i.e.*, the boundaries of the cells. We take the length of a cell to be d .

We shall assume small displacements of the atoms when a wave propagates along the lattice and also shall assume that all interactions are elastic. We shall not limit the distance at which interactions occur. The force on atom r in cell n due to atom s in cell $n + p$ is therefore

$$f_{n,r;n+p,s} = C_{prs}(y_{n+p,s} - y_{n,r}) \quad (17.1)$$

where C_{prs} is the interaction constant and is independent of n .

It follows that the force on particle s in cell $n + p$ due to particle r in cell n is

$$f_{n+p,s;n,r} = C_{-prs}(y_{n,r} - y_{n+p,s}) \quad (17.2)$$

According to Newton's third law

$$f_{n,r;n+p,s} = -f_{n+p,s;n,r} \quad (17.3)$$

and substituting Eqs. (17.1) and (17.2) into Eq. (17.3), we obtain

$$C_{prs}(y_{n+p,s} - y_{n,r}) = -C_{-psr}(y_{n,r} - y_{n+p,s})$$

or

$$C_{prs} = C_{-psr} \quad (17.4)$$

We take

$$C_{0rr} = 0 \quad (17.5)$$

since the term

$$C_{0rr}(y_{n,r} - y_{n,r}) = 0$$

and does not enter any of the calculations.

The total force acting on particle r in cell n will be given by

$$f_{n,r} = \sum_p \sum_s f_{n,r,n+p,s} = \sum_p \sum_s C_{prs}(y_{n+p,s} - y_{n,r}) \quad (17.6)$$

We assume a wave solution to Eq. (17.6) of the form

$$y_{n,r} = A_r e^{2\pi i(\nu t - ax)} \quad (17.7)$$

A_r is to be complex so as to contain the phase difference of particle r with particle 0, while x is the distance of the origin of the cell from the origin of the lattice.

$$x = nd$$

We may thus write Eq. (17.7) in the form

$$y_{n,r} = A_r e^{i(\omega t - kn)} \quad \left\{ \begin{array}{l} \omega = 2\pi\nu \\ k = 2\pi ad \end{array} \right\} \quad (17.8)$$

$y_{n,r}$ therefore has period $1/d$ in a and 2π in k as in Sec. 4. This means that k may be replaced by $k' = k + 2\pi p$ without affecting the solution. Substitution of Eq. (17.8) into Eq. (17.6) gives

$$\begin{aligned} f_{n,r} &= e^{i(\omega t - kn)} \sum_{p,s} C_{prs}(A_s e^{-ikp} - A_r) \\ &= M_r \frac{d^2 y_{n,r}}{dt^2} = -\omega^2 M_r A_r e^{i(\omega t - kn)} \end{aligned} \quad (17.9)$$

from which we obtain the following relation between ω and k :

$$\sum_s D_{r,s}(k) A_s = -\omega^2 M_r A_r \quad (17.10)$$

where the function $D_{r,s}(k)$ is defined by

$$\left. \begin{aligned} D_{r,s}(k) &= \sum_p C_{prs} e^{-ikp} & r \neq s \\ D_{r,r}(k) &= - \sum_{ps} C_{prs} + \sum_p C_{prp} e^{-ikp} \end{aligned} \right\} \quad (7.11)$$

The sum over s is to be taken over all atoms in a given cell and the sum over p is to be taken over all the cells.

The acoustical branch gives $A_r = A_s$ at $k = 0$, and hence

$$\begin{aligned} \omega^2 M_r &= \sum_s D_{r,s}(0) = \sum_{s \neq r} D_{r,s}(0) + D_{r,r}(0) \\ &= \sum_{s \neq r} \sum_p C_{prs} - \sum_{ps} C_{prs} + \sum_p C_{prp} \\ &= \sum_{s \neq r} \sum_p C_{prs} - \sum_{p, s \neq r} C_{prs} = 0 \\ \omega &= 0 \end{aligned}$$

For other values of k we write Eq. (17.10) as

$$\sum_s [D_{r,s}(k) + \omega^2 M_r \delta_{rs}] A_s = 0 \quad (17.12)$$

where δ is the Kronecker δ , defined by

$$\delta_{rs} = \begin{cases} 0 & r \neq s \\ 1 & r = s \end{cases}$$

Equation (17.12) gives N linear homogeneous equations for A_s , and the condition that they be consistent is that the determinant of the coefficients vanish; *i.e.*,

$$|D_{r,s}(k) + \omega^2 M_r \delta_{rs}| = 0 \quad (17.13)$$

Equation (17.13) is an equation of degree N in ω^2 , and hence there will be N values of ω^2 for a given k , *i.e.*, there will be N branches in the ω vs. k curve or the ν vs. a curve. One of these branches will be the acoustical branch, and the remaining $(N - 1)$ will be optical branches. ω^2 will be a periodic function of k since $D_{rs}(k)$ is a periodic function of k .

If we let $N \rightarrow \infty$, the number of optical branches becomes infinite, since we must have the total number of branches equal to the number of degrees of freedom of the system. The lattice will become a continuous string with some sort of periodic structure. We shall discuss the problem of the continuous periodic

string in a later chapter. If the string is continuous and uniform, ω is a linear function of k . Figure 17.2 shows the general appearance of the ω vs. k curves. The dotted curves are ω vs. k for a uniform continuous string.

The transition from the uniform continuous string to the continuous string with periodic structure (a loaded string, for

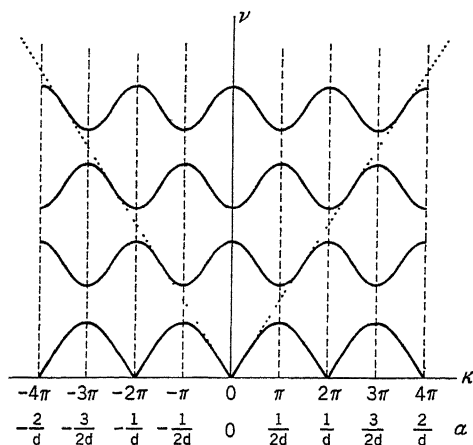


FIG. 17.2.

instance) is one of the problems of periodic perturbation sketched at the end of Sec. 16, and for the discussion of which the example of Sec. 16 will be used as a model. The change from the V-shaped dotted curve in Fig. 17.2 to the wavy curves occurs in a way similar to the change from the single dotted sine curve to the two solid curves in Fig. 16.1.

CHAPTER V

ENERGY VELOCITY, ENERGY FLOW, AND CHARACTERISTIC IMPEDANCE

13. General Discussion; Phase Velocity

So far we have discussed infinite lattices only. If we wish to apply our results to a finite lattice, we must add forces at the ends that will satisfy the boundary conditions. At the left end we must have a source of energy that will supply to the first particle the power that would have come to it if the lattice had extended indefinitely to the left. Then the propagation will depend on the frequency as noted at the end of Sec. 16. On the right end we must have a device that will absorb the energy that would have been absorbed by the omitted portion of the lattice extending indefinitely to the right. To set up the boundary conditions rigorously requires a discussion of the energy density, energy flow, and energy velocity in the lattice. This discussion will be carried on in the next few sections.

The one-dimensional mechanical lattice is an academic rather than a practical problem, and the only important instance of one-dimensional structures is found in electric lines, a discussion of which will be given in detail in the last chapters. It is, however, very useful to know how to set up the boundary conditions for the applications of the theory to two- and three-dimensional lattices. The method developed in this chapter is general and will be extended later to these problems, but it is easier to understand in the one-dimensional case.

The problems discussed are closely connected with the property of the structures of exhibiting *dispersion*. The wave velocity defined in the preceding chapters is known as *phase velocity*, since it is obtained from a comparison of the relative phase of the oscillations of two neighboring atoms. This phase velocity is the one to be used in formulas like

$$\lambda = V\tau, \quad V = \frac{v}{a}, \quad v = \frac{1}{\tau}, \quad a = \frac{1}{\lambda} \quad (18.1)$$

where τ is the period, ν the frequency, λ the wave length, and a the wave number. A *dispersive medium* is one for which the phase velocity V depends upon the frequency ν of oscillations. Many classical problems of wave propagation do not exhibit any such variation. Maxwell's equations of electromagnetism in vacuum lead to the equations of propagation of light and yield a constant velocity of propagation. Such is also the case for the standard equations for the propagation of sound waves, which result from a number of simplifications practically eliminating any frequency dependence of V . In such cases there is no difficulty in defining the velocity with which energy is transmitted through the medium by the wave motion. This velocity is simply equal to V . When, however, the transmitting medium is dispersive, the definition of energy velocity requires special attention and will be found to differ from phase velocity. This results from the fact that sine waves extending from $-\infty$ to $+\infty$ are the only waves to be transmitted without a change in their shape. Short signals or short impulses are distorted while they travel through the medium, and this distortion makes it difficult to define their average velocity. This is where the concept of *group velocity* comes in. A *group of waves*, or a *wave packet* (in the language of wave mechanics), is a signal of finite length, comprising only a limited number of wave lengths. We shall discuss the properties of such groups and the way in which they propagate through the medium and then compare the average velocity of the group with the energy velocity obtained from other definitions.

19. A Theorem from the Theory of Complex Variables

Following a method very commonly used, complex exponentials were introduced to represent waves or oscillations. For instance, the displacement and velocity of a particle in a wave were written

$$y = Ae^{2\pi i(\nu t - ax)}, \quad \dot{y} = \frac{dy}{dt} = 2\pi i\nu Ae^{2\pi i(\nu t - ax)}$$

respectively. Time derivatives are indicated by dots over the function. The order of the derivative is given by the number of dots.

$$\frac{dy}{dt} = \dot{y}, \quad \frac{d^2y}{dt^2} = \ddot{y}, \text{ etc.}$$

It must be recalled that such expressions should always be pre-

ceded by the sign R.P., meaning that only the *real part* of the quantity is taken into consideration, *viz.*,

$$y = A \cos 2\pi(\nu t - ax), \quad \dot{y} = -2\pi\nu A \sin 2\pi(\nu t - ax)$$

As long as we were working with equations linear in $y, \dot{y}, \ddot{y}, \dots$, we could drop the R.P. sign. But this is no longer allowed when double products or powers are encountered, since $y\dot{y}$, for instance, means

$$y\dot{y} = -2\pi\nu A^2 \cos 2\pi(\nu t - ax) \sin 2\pi(\nu t - ax) \neq 2\pi i\nu A^2 e^{4\pi i(\nu t - ax)}$$

This question will now be discussed.

We shall require the time average of the product of the real parts of complex functions on numerous occasions. There is a simple way of doing this by the following equation:

$$\overline{\text{R.P. } f \times \text{R.P. } F} = \frac{1}{2} \text{R.P. } (fF^*) \quad (19.1)$$

where f and F are complex functions of time of the form

$$f = f_0 e^{i(\omega t - \varphi)}, \quad F = F_0 e^{i(\omega t - \phi)} \quad (19.2)$$

The star means "complex conjugate of." Note that the time dependence of the two functions is the same.

We now prove that Eq. (19.1) is an identity.

$$\begin{aligned} \overline{\text{R.P. } f \times \text{R.P. } F} &= f_0 F_0 \overline{\cos(\omega t - \varphi) \cos(\omega t - \phi)} \\ &= f_0 F_0 \frac{1}{\tau} \int_0^\tau \cos(\omega t - \varphi) \cos(\omega t - \phi) dt, \end{aligned} \quad (19.3)$$

where τ is the period of f and F and is equal to $2\pi/\omega$. We may expand the integrand in Eq. (19.3) and obtain

$$\begin{aligned} \overline{\text{R.P. } f \times \text{R.P. } F} &= \frac{f_0 F_0}{\tau} \int_0^\tau \cos(\omega t - \varphi) \cos(\omega t - \varphi + \varphi - \phi) dt \\ &= \frac{f_0 F_0}{\tau} \int_0^\tau \cos(\omega t - \varphi) \\ &\quad [\cos(\omega t - \varphi) \cos(\varphi - \phi) - \sin(\omega t - \varphi) \sin(\varphi - \phi)] dt \end{aligned}$$

We replace $(\varphi - \phi)$ by ψ in the above so that

$$\overline{\text{R.P. } f \times \text{R.P. } \bar{F}} = \frac{f_0 F_0}{\tau} \int_0^\tau [\cos \psi \cos^2 (\omega t - \varphi) - \sin \psi \sin (\omega t - \varphi) \cos (\omega t - \varphi)] dt$$

The second term in the integrand becomes zero on integration, while the first term gives $\frac{\tau}{2} \cos \psi$. Therefore,

$$\overline{\text{R.P. } f \times \text{R.P. } \bar{F}} = \frac{1}{\tau} f_0 F_0 \frac{\tau}{2} \cos \psi = \frac{1}{2} f_0 F_0 \cos \psi = \frac{1}{2} \text{R.P. } (f F^*)$$

since

$$f F^* = f_0 F_0 e^{i(\omega t - \varphi)} e^{-i(\omega t - \phi)} = f_0 F_0 e^{i(\phi - \varphi)} = f_0 F_0 e^{-i\psi}$$

which proves the theorem.

20. Energy Density, Energy Flow, and Energy Velocity

First, we discuss the energy density and derive a mathematical expression for it. For the moment we shall confine ourselves to the *monatomic lattice with interactions between nearest neighbors only*. The theory of wave propagation in such a medium was discussed in Chap. III, Secs. 8 and 9.

We shall require the following relations, already derived in Eqs. (8.8), (8.10), and (9.2), for the discussion:

$$\left. \begin{aligned} y_n &= A e^{i(\omega t - kn)} & k &= 2\pi a d \\ \omega^2 &= \frac{2U''}{m} (1 - \cos k) = \frac{4U''}{m} \sin^2 \frac{k}{2} \\ W &= \text{phase velocity} = \frac{v}{a} = \frac{\omega d}{k} = W_\infty \left| \frac{\sin k/2}{k/2} \right| \end{aligned} \right\} \quad (20.1)$$

We shall temporarily drop the subscript on U'' .

The average energy density of the lattice will be the sum of the average potential-energy density and the average kinetic-energy density. The average potential-energy density is the average potential energy per cell divided by d , the length of the cell. Thus

$$\overline{E_{\text{pot}}} = \frac{1}{d} \text{R.P. } \frac{1}{2} U'' \overline{(y_n - y_{n-1})^2} \quad (20.2)$$

and since

$$y_n - y_{n-1} = \text{R.P. } A e^{i(\omega t - kn)} (1 - e^{ik})$$

it follows that

$$\begin{aligned}\overline{E_{\text{pot}}} &= \frac{U''}{2d} \frac{1}{2} \text{R.P. } (y_n - y_{n-1})(y_n - y_{n-1})^* \\ &= \frac{A^2 U''}{4d} (1 - e^{ik})(1 - e^{-ik}) = \frac{A^2 U''}{4d} (2 - e^{ik} - e^{-ik}) \\ &= \frac{A^2 U''}{4d} 2(1 - \cos k) = \frac{A^2 U''}{d} \sin^2 \frac{k}{2}\end{aligned}\quad (20.3)$$

The average kinetic-energy density is obtained in a similar manner; it is the average kinetic energy per cell divided by d , the length of the cell.

$$\overline{E_{\text{kin}}} = \text{R.P. } \frac{1}{d} \frac{1}{2} m \overline{(\dot{y}_n)^2} \quad (20.4)$$

and since

$$\dot{y}_n = \text{R.P. } \frac{dy_n}{dt} = \text{R.P. } A \frac{d}{dt} e^{i(\omega t - kn)} = \text{R.P. } i\omega y_n$$

we have

$$\overline{E_{\text{kin}}} = \text{R.P. } \frac{m}{2d} \overline{(i\omega y_n)^2} = \frac{\omega^2 m}{2d} \frac{1}{2} \text{R.P. } (iy_n)(iy_n)^* = \frac{m\omega^2 A^2}{4d} \quad (20.5)$$

Making use of the equation for ω^2 as a function of k in Eq. (20.1), we find that Eq. (20.5) reduces to

$$\overline{E_{\text{kin}}} = \frac{mA^2}{4d} \frac{4U''}{m} \sin^2 \frac{k}{2} = \frac{U'' A^2}{d} \sin^2 \frac{k}{2} = \overline{E_{\text{pot}}} \quad (20.6)$$

The total energy density therefore is

$$\overline{E} = \frac{2U'' A^2}{d} \sin^2 \frac{k}{2} \quad (20.7)$$

We shall need this relation later when we discuss the energy velocity.

The *energy flow* from one cell to the next will be the average power absorbed by the second cell from the first. With the first cell as cell n , this will be given by the negative product of the real part of the force $f_{n,n+1}$ on cell n due to cell $n+1$ and the real part of the velocity of the particle in cell n (for the case of the monatomic lattice). The negative product must be taken since $f_{n,n+1}$ is a force acting on particle n ; and hence the positive product would be the power furnished while the negative product would be the power absorbed by particle $n+1$. We have

$$\left. \begin{aligned} f_{n,n+1} &= U'''(y_{n+1} - y_n) \\ y_n &= A e^{i(\omega t - kn)} \\ \dot{y}_n &= i\omega A e^{i(\omega t - kn)} = i\omega y_n \end{aligned} \right\} \quad (20.8)$$

The average power Φ (time average) absorbed by cell $n + 1$ is thus

$$\begin{aligned} \Phi &= -\overline{\text{R.P. } f_{n,n+1} \times \text{R.P. } \dot{y}_n} = -\frac{1}{2} \text{R.P. } (f_{n,n+1} \dot{y}_n^*) \\ &= -\frac{U''' A^2}{2} \text{R.P. } (e^{-i\omega t} - 1)(i\omega)^* \\ &= \frac{U''' A^2 \omega}{2} \text{R.P. } [\sin k - i(1 - \cos k)] = \frac{U''' A^2 \omega}{2} \sin k \quad (20.9) \end{aligned}$$

Substituting the value for ω given in Eq. (20.1),

$$\Phi = U''' A^2 \sqrt{\frac{U'''}{m}} \sin \frac{k}{2} \sin k \quad (20.10)$$

The energy flow gives us the energy passing from cell n to cell $(n + 1)$ per unit time. A quantity closely connected with this is the *energy velocity*. It is defined as the energy flow divided by the energy density and gives the rate at which energy flows along the lattice. We denote the energy velocity by U_e .

$$U_e = \frac{\Phi}{E} = \frac{U''' A^2 \sqrt{\frac{U'''}{m}} \sin \frac{k}{2} \sin k}{2 \frac{U''}{d} A^2 \sin^2 \frac{k}{2}} = d \sqrt{\frac{U'''}{m}} \cos \frac{k}{2} \quad (20.11)$$

The energy velocity can always be defined, even if absorption is present. The meaning of Eq. (20.11) will appear clearly if it is compared with the formula giving the flow of matter in a fluid: let ρ be the density of the fluid and v its velocity. The flow of matter is $\Phi = \rho v$; hence the ratio Φ/ρ is the velocity of the fluid. In a similar way the ratio Φ/E of energy flow to energy density obviously yields a velocity that is the velocity with which energy is flowing through the system. More detailed explanations and examples can be found in a report by the author.¹

21. Group Velocity and Propagation of a Signal

Having explained in Sec. 18 the meaning attached to the expression "group" or "wave packet," we may immediately

¹ "Congrès international d'électricité," Paris, 1932, vol. II, p. 739.

proceed to the discussion of the simplest example, obtained by considering the wave motion due to the superposition of two sine waves of frequencies $\nu_0 \pm \Delta\nu$ and equal amplitudes A . The wave $\nu_0 + \Delta\nu$ has wave number $a_0 + \Delta a$, while $\nu_0 - \Delta\nu$ has wave number $a_0 - \Delta a$. Thus the wave $\nu_0 - \Delta\nu$ has the equation of motion

$$y_- = A \cos 2\pi[(\nu_0 - \Delta\nu)t - (a_0 - \Delta a)x]$$

while the equation for the wave $\nu_0 + \Delta\nu$ is

$$y_+ = A \cos 2\pi[(\nu_0 + \Delta\nu)t - (a_0 + \Delta a)x]$$

To obtain the resultant motion, we add the two disturbances algebraically.

$$\begin{aligned} y &= y_- + y_+ = A \{ \cos 2\pi[(\nu_0 - \Delta\nu)t - (a_0 - \Delta a)x] \\ &\quad + \cos 2\pi[(\nu_0 + \Delta\nu)t - (a_0 + \Delta a)x] \} \\ &= 2A \cos 2\pi(\nu_0 t - a_0 x) \cos 2\pi(\Delta\nu \cdot t - \Delta a \cdot x) \quad (21.1) \end{aligned}$$

This represents a *modulated wave* with an average frequency ν_0 in the *carrier wave*

$$\cos 2\pi(\nu_0 t - a_0 x) \quad (21.2)$$

and a slowly variable amplitude considered as the *modulation*

$$A \cos 2\pi(\Delta\nu \cdot t - \Delta a \cdot x) \quad (21.3)$$

The *phase velocity* of the carrier wave is

$$V = \frac{\nu_0}{a_0} \quad (21.4)$$

In the same way the modulation is seen to move with a velocity given by $\Delta\nu/\Delta a$. In the limit when the two frequencies become equal,

$$U_g = \text{group velocity} = \frac{\partial \nu}{\partial a} \quad (21.5)$$

There is no difficulty in defining U_g as long as the medium is purely dispersive, *i.e.*, $\nu = \nu(a)$; but if absorption also occurs, a becomes complex or imaginary and the group velocity ceases to have a clear physical meaning.

So far we have assumed zero coefficient of absorption in the monatomic lattice. Therefore,

$$\begin{aligned}
 U_g &= \frac{\partial \nu}{\partial a} = d \frac{\partial \omega}{\partial k} = d \frac{\partial}{\partial k} \left(2 \sqrt{\frac{U''}{m}} \sin \frac{k}{2} \right) \\
 &= 2d \sqrt{\frac{U''}{m}} \frac{1}{2} \cos \frac{k}{2} = d \sqrt{\frac{U''}{m}} \cos \frac{k}{2} = U_e \quad (21.6)
 \end{aligned}$$

Thus if no absorption is present, the group velocity and the energy velocity are the same. However, the group velocity breaks down for cases with absorption, while the energy velocity can always be defined.

The motion represented by Eq. (21.1) is best described in the following way: It consists of a succession of wavelets (21.2) of frequency ν_0 and wave length $1/\lambda_0$. At a certain instant of time the average amplitude of these wavelets is given by the modulation (21.3). If we do not pay attention to the detailed motion of

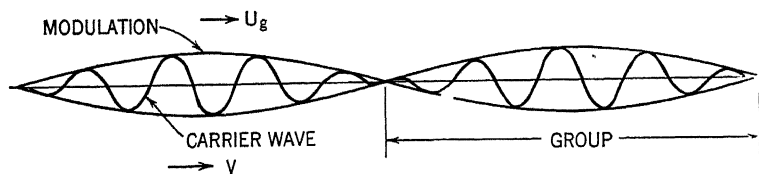


FIG. 21.1.

the wavelets and look only at the average amplitude distribution, we see this amplitude curve (21.3) move forward with the group velocity U_g . But if we look at the phenomenon more carefully, we notice the wavelets moving inside the envelope (21.3) with their own phase velocity V (Fig. 21.1). A well-known example of such an appearance is found when surface waves are created by throwing a stone into a pond. The preceding example is just one among many similar ones, and the results obtained are to a large extent independent of the shape of the group or of the type of the modulation curve. It is characterized by the following feature: The modulation curve propagates without distortion and exhibits a well-defined velocity. The absence of distortion is obviously connected with the *absence of attenuation*. In an absorbing medium with attenuation the definition of a group velocity loses its accuracy.

Furthermore, the absence of distortion can be obtained only if the wave packet results from the superposition of elementary waves whose frequencies lie within a small interval. In the preceding example we had just two frequencies $\nu_0 + \Delta\nu$ and

$\nu_0 - \Delta\nu$ within a finite interval $2\Delta\nu$. We should find similar results with wave groups obtained by superposition of any number of waves with frequencies within a given interval $\Delta\nu$. In other words, the Fourier analysis of the group must yield a spectrum of finite length, which in the limit can be made infinitely small and can allow for the transition from $\Delta\nu/\Delta a$ to the derivative $\partial\nu/\partial a$ of Eq. (21.5).

There are other types of groups or signals whose Fourier spectra extend from $-\infty$ to $+\infty$ in the frequency range. For such signals it is impossible to go to the limit $\Delta\nu \rightarrow 0$, and the definition of a group velocity again loses its accuracy. This results in the fact that the modulation curve progressively changes its shape in the course of propagation and is more and more distorted as time goes on. These general remarks will be illustrated in a few precise examples, where we shall use some well-known formulas involving Fourier integrals. Let $C(t)$ be an even function of time and $B(\nu)$ its frequency spectrum.

$$C(t) = C(-t)$$

Then the Fourier transformation reads

$$\left. \begin{aligned} C(t) &= \int_{-\infty}^{\infty} B(\nu) \cos 2\pi\nu t \, d\nu \\ B(\nu) &= \int_{-\infty}^{\infty} C(t) \cos 2\pi\nu t \, dt \end{aligned} \right\} \quad (21.7)$$

The last formula obviously yields B as an even function.

$$B(\nu) = B(-\nu)$$

and the reciprocity between C and B results in the following statement: If a signal $C(t)$ has a spectrum $B(\nu)$, then a reciprocal signal $B(t)$ will be represented by a spectrum $C(\nu)$. We may use the signal $C(t)$ as a modulation curve on a carrier oscillation of frequency ν_0 , and we obtain a new even function

$$C_1(t) = C(t) \cos 2\pi\nu_0 t \quad (21.8)$$

The frequency spectrum of C_1 is easily obtained.

$$\begin{aligned} B_1(\nu) &= \int_{-\infty}^{\infty} C_1(t) \cos 2\pi\nu t \, dt = \int_{-\infty}^{\infty} C(t) \cos 2\pi\nu t \cos 2\pi\nu_0 t \, dt \\ &= \frac{1}{2} \int_{-\infty}^{\infty} C(t) [\cos 2\pi(\nu_0 + \nu)t + \cos 2\pi(\nu_0 - \nu)t] \, dt \\ &= \frac{1}{2} [B(\nu_0 + \nu) + B(\nu_0 - \nu)] = \frac{1}{2} [B(\nu + \nu_0) + B(\nu - \nu_0)]. \end{aligned} \quad (21.9)$$

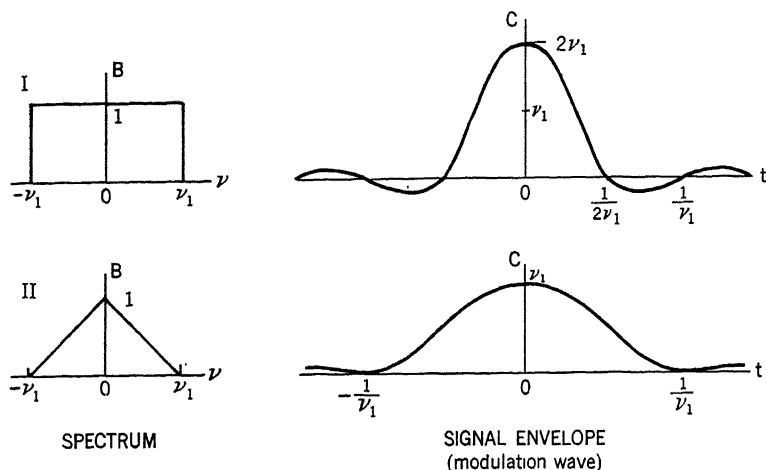


FIG. 21.2.

$B_1(\nu)$ is again an even function of ν . The spectrum $B(\nu)$ of the original signal is centered on the origin O . The new spectrum is obtained by the average of two such curves, translated by $\pm \nu_0$.

Let us give a few examples of Fourier transformations, corresponding to the reciprocal curves of Fig. 21.2.

Spectrum $B(\nu)$

Signal $C(t)$

I. Rectangular

$$B(\nu) = \begin{cases} 1 & |\nu| < \nu_1 \\ 0 & |\nu| > \nu_1 \end{cases} \quad C(t) = 2\nu_1 \frac{\sin 2\pi\nu_1 t}{2\pi\nu_1 t} \quad (21.10)$$

II. Triangular

$$B(\nu) = \begin{cases} 1 - \frac{|\nu|}{\nu_1} & |\nu| < \nu_1 \\ 0 & |\nu| > \nu_1 \end{cases} \quad C(t) = \nu_1 \left(\frac{\sin \pi\nu_1 t}{\pi\nu_1 t} \right)^2 \quad (21.11)$$

Both signals exhibit a finite spectrum, while the signals themselves extend from $t = -\infty$ to $t = \infty$ with a strong maximum at $t = 0$. The reciprocal signals would be finite signals (rectangular or triangular) $B(t)$ with infinite spectra $C(\nu)$.

We now want to prove that a signal with a finite spectrum propagates in a way similar to the beats of Eq. (21.1). Let us take a $C(t)$ modulation impressed upon a ν_0 carrier, as in Eq. (21.8). We assume this motion to be impressed on the atom at

$x = 0$, and we compute the motion at a distance x from the origin. This means only replacing

$$\nu t \quad \text{by} \quad \nu t - a(\nu)x$$

where $a(\nu)$ is the wave number $1/\lambda$ as a function of ν for the transmitting medium. Taking our modulated signal (21.8), we obtain according to Eq. (21.9)

$$\begin{aligned} C_1(t) &= \int_{-\infty}^{\infty} B_1(\nu) \cos 2\pi\nu t \, d\nu = 2 \int_0^{\infty} B_1 \cos 2\pi\nu t \, d\nu \\ &= \int_0^{\infty} [B(\nu + \nu_0) + B(\nu - \nu_0)] \cos 2\pi\nu t \, d\nu \quad (21.12) \end{aligned}$$

Let us write

$$\nu = \nu_0 + \mu \qquad \nu_0 > \nu_1$$

Then, for both examples (21.10) and (21.11), the first term in $B(\nu + \nu_0)$ is always zero, and we find

$$C_1(t) = \int_{\mu=-\nu_1}^{\mu=+\nu_1} B(\mu) \cos 2\pi(\nu_0 + \mu)t \, d\mu \quad (21.13)$$

This is the motion of the point at the origin $x = 0$. For a point at distance x we obtain

$$C_1(t, x) = \int_{-\nu_1}^{\nu_1} B(\mu) \cos 2\pi[(\nu_0 + \mu)t - a(\nu_0 + \mu)x] \, d\mu \quad (21.14)$$

But ν_1 is supposed to be small enough to allow for an expansion.

$$a(\nu_0 + \mu) = a(\nu_0) + \mu \left(\frac{\partial a}{\partial \nu} \right)_0 \quad |\mu| < \nu_1 \quad (21.15)$$

Hence

$$(\nu_0 + \mu)t - ax = \nu_0 t - a_0 x + \mu \left(t - \frac{\partial a}{\partial \nu_0} x \right)$$

Expanding the cosine in Eq. (21.14) and recalling that B is even, we obtain

$$\begin{aligned} C_1(t, x) &= \cos 2\pi(\nu_0 t - a_0 x) \int_{-\nu_1}^{\nu_1} B(\mu) \cos \mu \left(t - \frac{\partial a}{\partial \nu_0} x \right) \, d\mu \\ &= \cos 2\pi(\nu_0 t - a_0 x) C \left(t - \frac{\partial a}{\partial \nu_0} x \right) \quad (21.16) \end{aligned}$$

This is the result announced: Individual wavelets propagate with their own phase velocity as shown by the cosine term. The

modulation curve $C\left(t - \frac{\partial a}{\partial \nu_0} x\right)$ moves along without distortion with *group velocity* (21.5).

The proof rests upon the assumption of a finite spectrum whose limit ν_1 can be taken small enough to use Eq. (21.15). Signals with an infinite spectrum are always more or less distorted.

The results are plotted on the curves of Fig. 21.3. The upper curve is the familiar one of ν against a , as in Fig. 2.4, for the monatomic lattice with interactions between nearest neighbors only.

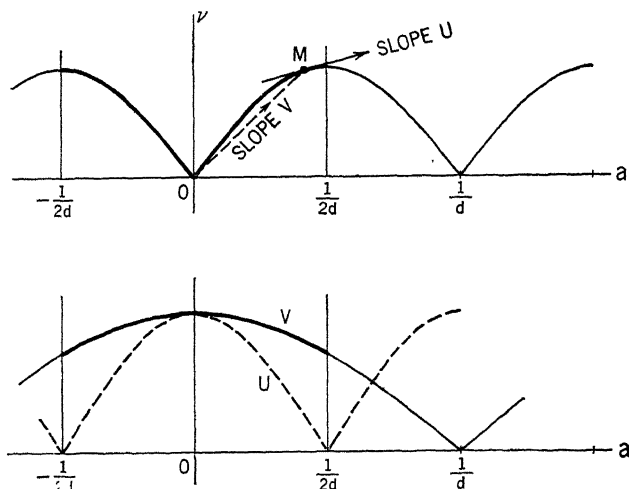


FIG. 21.3.

If a point M is taken on the curve, the absolute value of the slope of a chord OM gives the phase velocity $V = \nu/a$, while the tangent at M yields the group velocity $U = \partial \nu / \partial a$. Curves for V and U as functions of a are given at the bottom. The V curve does not exhibit any singularity at $a = 1/2d$ (cf. discussion of Fig. 2.2), but the group velocity U drops to zero on the limit $\pm 1/2d$ of the interval, a feature that checks very well with our description of these limit waves as standing waves (Secs. 2, 4, and 9).

22. Preliminary Definition of Characteristic Impedance

Impedance for a mechanical system is defined as the ratio of force exerted to velocity. In discussing periodic lines, *i.e.*, lattice structures with cells that repeat themselves periodically or a

continuous line with a periodic structure, we mean by *characteristic impedance* the ratio between force and velocity for a single sine wave at the entrances or exits of the cells. For an infinite line, along which a single wave is propagated, this should be the same for all cells. The same is true if the line is finite but has been provided with suitable forces for absorbing and furnishing energy, so that, except at the ends, the line behaves as though it were infinite. Obviously, a determination of characteristic impedance gives us the impedances that must terminate a finite line if it is to behave as an infinite line exhibiting no reflection.

In this section we shall consider only the lattice consisting of like particles with interactions between nearest neighbors only. Then the characteristic impedance is given by

$$f_{n,n+1} = -Z\dot{y}_n \quad (22.1)$$

where $f_{n,n+1}$ is the force exerted by particle $n+1$ on particle n , \dot{y}_n its velocity, and Z the impedance. From the earlier discussion of the problem [cf. Eqs. (20.1) and (20.8)]

$$\begin{aligned} y_n &= A e^{i(\omega t - kn)} \\ f_{n,n+1} &= U'''(y_{n+1} - y_n) = U'''y_n(e^{-ik} - 1) \\ \dot{y}_n &= i\omega y_n \end{aligned}$$

and therefore

$$Zi\omega = U'''(1 - e^{-ik}) = U'''(1 - \cos k + i \sin k) \quad (22.2)$$

We allow Z to be complex and set

$$Z = Z_r + iZ_i \quad (22.3)$$

Equating real and imaginary parts in Eq. (22.2), we obtain

$$\left. \begin{aligned} Z_i &= \frac{U'''}{\omega} (\cos k - 1) = -\frac{2U'''}{\omega} \sin^2 \frac{k}{2} \\ &= -\sqrt{U'''m} \sin \frac{k}{2} \\ Z_r &= \frac{U'''}{\omega} \sin k = \frac{2U'''}{\omega} \sin \frac{k}{2} \cos \frac{k}{2} = \sqrt{U'''m} \cos \frac{k}{2} \end{aligned} \right\} \quad (22.4)$$

on making use of $\omega = 2\sqrt{U'''/m} \sin k/2$. With the results of the last section,

$$Z_r = \frac{m}{d} U_e \quad U_e, \text{ energy velocity} \quad (22.5)$$

We can interpret Z_r and Z_i completely by replacing

$$U''(1 - \cos k)$$

by C and noting that C is always positive. Further, Z_r is positive for the fundamental interval $-\pi < k < \pi$, since $\cos k/2$ is positive in this region.

$$\begin{aligned} f_{n,n+1} &= -Z\dot{y}_n = -Z_r\dot{y}_n - iZ_i\dot{y}_n = -Z_r\dot{y}_n + i\frac{C}{\omega}\dot{y}_n \\ &= -Z_r\dot{y}_n - \frac{C}{i\omega}\dot{y}_n = -Z_r\dot{y}_n - Cy_n \quad (22.6) \end{aligned}$$

We have now split the force acting on mass n into two parts: the first term gives a viscous force and the second an elastic force.

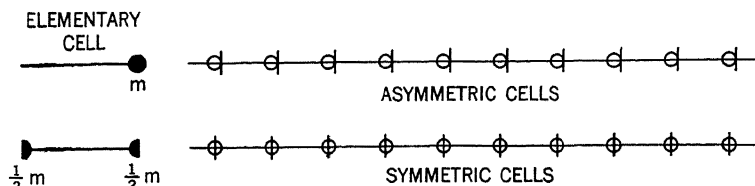


FIG. 22.1.

There is a disadvantage to the treatment we have just given; we have taken the cells to contain a whole particle at one end. If we try to find the impedance at the other end of the cell, we run into a difficulty because there is nothing there to exert a force or to have a velocity, since the cell must contain exactly one period of the lattice. The cells, as we have chosen them, are asymmetric and do not lend themselves conveniently to impedance considerations.

The difficulty may be obviated by defining the cells differently and making them symmetric. We take the cells to be of length d (where d is the distance between the particles) as before, but we associate with the cell half of each of the masses at the ends. This makes the cell symmetric. The symmetric and asymmetric cells are shown in Fig. 22.1, where the vertical lines denote the boundaries of the cells in the two cases. Figure 22.2 shows a line composed of symmetric cells and terminating on an impedance Z_s .

Let us find this impedance Z_s at the right end of the cell containing half of mass $n - 1$ at the left and half of mass n at the right end. The force on mass n due to mass $n - 1$ is given by

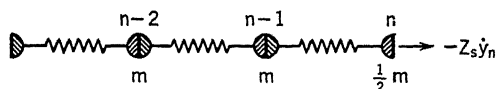
$U''(y_{n-1} - y_n)$, and the impedance term is $-Z_s \dot{y}_n$. Therefore, the equation of motion at the right end $\frac{1}{2}m$ of the cell will be

$$U''(y_{n-1} - y_n) - Z_s \dot{y}_n = \frac{1}{2}m \ddot{y}_n \quad (22.7)$$

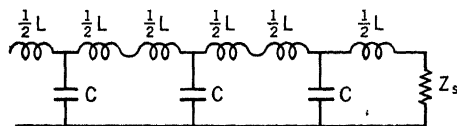
The force due to mass $n + 1$ is thought of as acting on the half of mass n in the next cell to the right. If that force were added to this, we would find exactly the equation of motion obtained before; the impedance would have opposite signs for the two halves of mass n and would cancel, so that the solution for y_n would be the same as before. Substituting the solution in Eq. (22.7) gives an equation for Z_s .

$$U''(e^{ik} - 1) - i\omega Z_s = \frac{1}{2}m(i\omega)^2 = -\frac{1}{2}m\omega^2 \quad (22.8)$$

We want to show that Z_s gives rise only to a viscous force in this model, and that the elastic force has already been explicitly



A LINE WITH SYMMETRIC CELLS AND
TERMINATION ON A CHARACTERISTIC IMPEDANCE Z_s



ELECTRICAL ANALOGUE

FIG. 22.2.

included; *i.e.*, we now have Z_s real. Equating the real and imaginary parts of Eq. (22.8), we obtain

$$\left. \begin{aligned} Z_s &= \frac{U''}{\omega} \sin k = \sqrt{U''m} \cos \frac{k}{2} = Z_r \\ 2 \frac{U''}{m} (1 - \cos k) &= 4 \frac{U''}{m} \sin^2 \frac{k}{2} = \omega^2 \end{aligned} \right\} \quad (22.9)$$

The impedance at the other end of the cell is easily obtained in exactly the same manner. The equation of motion will be

$$U''(y_n - y_{n-1}) + Z_s \dot{y}_{n-1} = \frac{1}{2}m \ddot{y}_{n-1}$$

and Eq. (22.8) will be replaced by

$$U''(e^{-ik} - 1) + i\omega Z_s = -\frac{1}{2}m\omega^2$$

from which we obtain

$$\left. \begin{aligned} Z_s &= \frac{U''}{\omega} \sin k = Z_r \\ \omega^2 &= 4 \frac{U''}{m} \sin^2 \frac{k}{2} \end{aligned} \right\} \quad (22.10)$$

Summarizing the whole discussion, we have found that if we wish to end the line with particle n (mass m) on the right we must use asymmetric cells and apply an impedance $Z_r + iZ_i$, where Z_r is the viscous force and Z_i is an elastic force. If such an impedance is applied, the remaining masses will vibrate exactly as if the lattice were infinite. If we wished to make mass n the last mass on the left, we should have to take asymmetric cells with masses on the left instead of on the right end of the cells.

On the other hand, we may take symmetric cells and take particle n with $\frac{1}{2}m$ to be the end of the lattice, either on the right or on the left. The remaining masses will vibrate as if the line were infinite if we apply only a viscous force with impedance Z_s to the remaining half of mass n ; and the elastic force Z_i occurring in the case of asymmetric cells is automatically taken care of by removing half of the terminating mass.

We may make a few remarks on the low-pass electric line of Fig. 22.2. Using the classical analogy

$$M \sim L, \quad C \sim \frac{1}{U''}$$

we obtain

$$Z_s = \sqrt{\frac{L}{C}} \cos \frac{k}{2} \quad (22.11)$$

or, for infinite wave length,

$$Z_s = \sqrt{\frac{L}{C}}$$

L and C are, of course, to be taken as inductance and capacity per cell, respectively, though inductance and capacity per unit length of line give the same result since the ratio is the quantity occurring. The impedance of electric lines will be discussed in detail later.

The *energy flow* through the lattice was defined in Sec. 20, Eq. (20.9), by the formula

$$\Phi = -\overline{\text{R.P. } f_{n,n+1} \times \text{R.P. } \dot{y}_n^*} = -\frac{1}{2} \text{R.P. } (f_{n,n+1} \dot{y}_n^*)$$

and now we obtain the characteristic impedance from Eq. (22.1).

$$f_{n,n+1} = -Z \dot{y}_n$$

Comparing these formulas, we find

$$\Phi = \frac{1}{2} \text{R.P. } (Z \dot{y}_n \dot{y}_n^*) = \frac{1}{2} Z_r |\dot{y}_n|^2 \quad (22.12)$$

This yields a *relation* between *energy flow* and the *real part* of the *characteristic impedance*, and it must be emphasized here that this real part Z_r is the one upon which all our different definitions agree simultaneously.

At the limit of indefinitely long wave lengths, our formulas will reduce to the well-known ones for a continuous structure. We have already discussed this transition in Sec. 9 and obtained

$$\epsilon = dU''_1 \quad \text{elasticity modulus} \quad (9.3)$$

$$\rho = \frac{m}{d} \quad \text{density}$$

$$V_\infty = d \sqrt{\frac{U''_1}{m}} = \sqrt{\frac{\epsilon}{\rho}} \quad \text{phase velocity} \quad (9.3a)$$

In the same way we find now

$$\left. \begin{aligned} U_\infty &= V_\infty \\ Z_r &= \sqrt{m U''_1} = \sqrt{\epsilon \rho} \end{aligned} \right\} \quad (22.13)$$

When there is no dispersion and V is a constant, then, of course,

$$U = \frac{\partial \nu}{\partial a} = \frac{\nu}{a} = V$$

The group velocity is equal to the phase velocity. The formula (22.13) for the characteristic impedance is the usual one.

23. Junction of Two Lattices

We are now in a position to discuss the behavior of waves at the junction of two monatomic lattices with the particles spaced at distance d from one another and with interactions between nearest neighbors only. Let us suppose that we have two such lattices with phase velocities W_1 and W_2 , group velocities U_1 and

U_2 , energy flows Φ_1 and Φ_2 , impedances Z_1 and Z_2 , and amplitudes A_1 and A_2 at the ends to be joined. We shall call the ratio of the amplitudes the transformer ratio T .

$$T = \frac{A_2}{A_1} \quad \text{or} \quad A_2 = TA_1 \quad (23.1)$$

We have obtained a relation between energy flow and characteristic impedance for any lattice [Eq. (22.12)]. If we substitute the value of \dot{y}_n and take the time average in the usual fashion, Eq. (22.12) becomes

$$\Phi = \frac{1}{2} Z_r \omega^2 A^2 \quad (23.2)$$

Later we shall find it convenient to use Eq. (23.2) as the defining equation for the characteristic impedance.

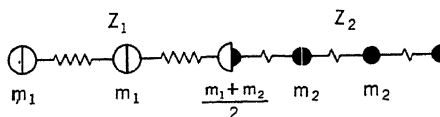


FIG. 23.1.

If we now join the two lines described above and require that the ends to be joined be the ends of cells in the two lattices, we obtain

$$\Phi_1 = \frac{1}{2} Z_{r1} \omega_1^2 A_1^2 \quad (23.3)$$

$$\Phi_2 = \frac{1}{2} Z_{r2} \omega_2^2 A_2^2 \quad (23.4)$$

In general, we shall take $\omega_1 = \omega_2$ and the condition that there be no energy loss and no reflected wave is

$$\Phi_1 = \Phi_2 \quad \text{or} \quad Z_{r1} A_1^2 = Z_{r2} A_2^2 \quad \text{or} \quad \frac{Z_{r1}}{Z_{r2}} = T^2 \quad (23.5)$$

T may be easily computed for two monatomic lattices when the conditions at the junction are specified. Let us take, for instance, the lattices as divided into symmetrical cells; then Z_1 and Z_2 are real. At the junction in this structure we have a particle, one half of which belongs to one lattice and the other half of which belongs to the other lattice (see Fig. 23.1). Since the two halves must move together, we have the condition

$$A_1 = A_2 \quad \text{or} \quad T = 1 \quad (23.6)$$

Hence, the condition for zero energy loss and no reflection at the junction is

$$Z_1 = Z_2 \quad (23.7)$$

We obtained the expression for the characteristic impedance in the last section. If the masses in the two lattices are different, it follows that the elastic coefficients must also be different.

In general, *reflection* occurs at the junction of two lattices, and the coefficient of reflection may be obtained in terms of the constants of the lattices. If reflection occurs, we have three waves at the junction: the incident wave with amplitude A_1 , the transmitted wave with amplitude A_2 , and the reflected wave with amplitude A_3 . The energy flows associated with these waves are given by

$$\left. \begin{aligned} \Phi_1 &= \frac{1}{2} Z_{r_1} \omega^2 A_1^2 \\ \Phi_2 &= \frac{1}{2} Z_{r_2} \omega^2 A_2^2 \\ \Phi_3 &= \frac{1}{2} Z_{r_1} \omega^2 A_3^2 \end{aligned} \right\} \quad (23.8)$$

since the incident and reflected waves propagate in the first lattice and the transmitted wave in the second lattice. We wish to obtain the coefficient of reflection

$$R = \frac{A_3}{A_1} \quad (23.9)$$

To do this, we note that we have two conditions for the expressions in Eq. (23.8). Conservation of energy requires that

$$\Phi_1 = \Phi_2 + \Phi_3 \quad (23.10)$$

and the condition that the two halves of the particle at the junction move together is

$$A_2 = A_1 + A_3 \quad (23.11)$$

since the resultant of the motion due to the incident and reflected waves is an algebraic sum of these waves.

From Eq. (23.10) we obtain

$$Z_{r_1} A_1^2 = Z_{r_2} A_2^2 + Z_{r_1} A_3^2$$

or

$$\frac{Z_{r_2}}{Z_{r_1}} \left(\frac{A_2}{A_1} \right)^2 = 1 - \left(\frac{A_3}{A_1} \right)^2$$

or

$$\left(\frac{A_2}{A_1} \right)^2 = \frac{Z_{r_1}}{Z_{r_2}} \left[1 - \left(\frac{A_3}{A_1} \right)^2 \right] \quad (23.12)$$

and from Eq. (23.11)

$$\left(\frac{A_2}{A_1}\right)^2 = 1 + 2\frac{A_3}{A_1} + \left(\frac{A_3}{A_1}\right)^2 \quad (23.13)$$

Combining Eq. (23.12) with Eq. (23.13), we obtain a quadratic equation in $(A_3/A_1) = R$.

$$\left(\frac{A_3}{A_1}\right)^2 \left(1 + \frac{Z_{r1}}{Z_{r2}}\right) + 2\left(\frac{A_3}{A_1}\right) + 1 - \frac{Z_{r1}}{Z_{r2}} = 0$$

or

$$R^2 \left(1 + \frac{Z_{r1}}{Z_{r2}}\right) + 2R + 1 - \frac{Z_{r1}}{Z_{r2}} = 0 \quad (23.14)$$

$$(R + 1)^2 + \frac{Z_{r1}}{Z_{r2}}(R^2 - 1) = 0$$

$$(R + 1) \left[R + 1 + \frac{Z_{r1}}{Z_{r2}}(R - 1) \right] = 0 \quad (23.15)$$

The solution

$$R = -1 \quad (23.16)$$

is trivial. The incident and reflected waves are just out of phase, and the particle at the junction is at rest. The transmitted wave has amplitude zero, as may be seen by substitution of Eq. (23.16) in Eqs. (23.11) to (23.13). The other solution is

$$R = \frac{Z_{r1} - Z_{r2}}{Z_{r1} + Z_{r2}} \quad (23.17)$$

which gives the coefficient of reflection for amplitudes. The coefficient of reflection for intensities is

$$R' = |R|^2$$

The coefficient of transmission T' for intensities and the transformation ratio T are given by

$$T = \frac{A_2}{A_1} = \text{transformation ratio} = 1 + R = \frac{2Z_{r1}}{Z_{r1} + Z_{r2}}$$

$$\begin{aligned} T' = \text{coefficient of transmission} &= |T|^2 = \frac{Z_{r1}}{Z_{r2}}(1 - R^2) \\ &= \frac{4Z_{r1}^2}{(Z_{r1} + Z_{r2})^2} \end{aligned}$$

24. General Definitions of Characteristic Impedance

So far we have discussed the monatomic lattice with interactions between nearest neighbors only. We may extend the

treatment to the case of a monatomic lattice with the range of interaction unlimited and use the definitions and formulas of Secs. 8 and 10. The potential energy was obtained in Eq. (8.5).

$$U = U_0 + \sum_n \sum_{p>0} \left[(y_{n+p} - y_n) U'_p + \frac{1}{2} (y_{n+p} - y_n)^2 U''_p \right] \quad (8.5)$$

The *potential-energy density* is the energy per cell divided by d . In a single wave, each particle has a sinusoidal motion and $\overline{(y_{n+p} - y_n)^t}$ is zero; hence

$$\overline{E_{\text{pot}}}^t = \frac{1}{2d} \sum_{p>0} \overline{U''_p (y_n - y_{n+p})^2}^t \quad (24.1)$$

A single wave $y_n = A e^{i(\omega t - kn)}$ propagating to the right gives

$$\begin{aligned} \overline{(y_n - y_{n+p})^2}^t &= \frac{1}{2} \text{R.P. } (y_n - y_{n+p})(y_n - y_{n+p})^* \\ &= \frac{1}{2} A^2 \text{R.P. } (1 - e^{-ikp})(1 - e^{ikp}) \\ &= A^2 (1 - \cos kp) \end{aligned} \quad (24.2)$$

and hence

$$\overline{E_{\text{pot}}}^t = \frac{A^2}{2d} \sum_{p>0} U''_p (1 - \cos kp) \quad (24.3)$$

The *average kinetic-energy density* is the same as before [Eq. (20.5)]:

$$\overline{E_{\text{kin}}}^t = \frac{m}{4d} A^2 \omega^2 = \frac{A^2}{2d} \sum_{p>0} U''_p (1 - \cos kp) \quad (24.4)$$

on substitution for ω^2 [Eq. (10.1)]. Thus

$$\overline{E_{\text{kin}}}^t = \overline{E_{\text{pot}}}^t = \frac{1}{2} \overline{E}^t = \frac{A^2}{2d} \sum_{p>0} U''_p (1 - \cos kp) \quad (24.5)$$

The larger range of interaction complicates the problem of finding the *energy flow*. Let us compute the flow of energy to the right from all of the cells to the left of a certain particle that can be taken at the origin ($n = 0$). To do this, we must compute the force exerted by a particle $n < 0$ on all particles interacting with it on the right. The force on particle n due to particle $n + p$ is $f_{n,n+p}$ in the previous notation. The subscripts denoting par-

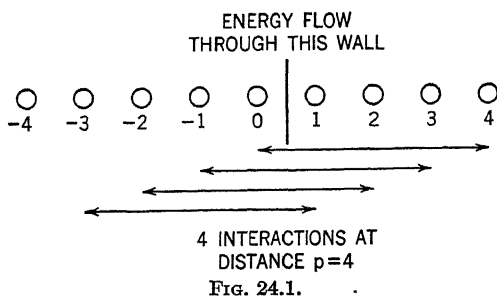
ticle numbers have been omitted since each cell contains essentially only one particle.

$$f_{n,n+p} = U''_p(y_{n+p} - y_n)$$

and the average energy flow due to this force will be given by

$$-\overline{f_{n,n+p}\dot{y}_n^t} = -\frac{1}{2} \text{R.P.} (f_{n,n+p}\dot{y}_n^*) \quad (24.6)$$

Again the minus sign occurs since $f_{n,n+p}$ is the force acting on particle n due to particle $n+p$ rather than the force exerted by particle n on particle $n+p$. The latter is, of course, the negative of the former. The right-hand side of Eq. (24.6) may be



given explicitly by substituting the well-known exponential expressions for y_n and y_{n+p} .

$$\begin{aligned} \frac{1}{2} \text{R.P.} (f_{n,n+p}\dot{y}_n^*) &= \frac{A^2 U''_p}{2} \text{R.P.} (e^{-ikp} - 1)(i\omega)^* \\ &= -\frac{1}{2} U''_p A^2 \omega \sin kp \end{aligned}$$

which gives for Eq. (24.6)

$$-\overline{f_{n,n+p}\dot{y}_n^t} = \frac{1}{2} U''_p A^2 \omega \sin kp \quad (24.7)$$

There will be p terms of the type (24.7) contributing to the energy flow across particle $n = 0$, since each of the particles $n = 0, -1, -2, \dots, -p+1$ will furnish this amount of power to the first p particles to the right of particle 0 (see Fig. 24.1). Thus the total *energy flow* will be given by the following sum over p :

$$\Phi = \sum_p \frac{1}{2} U''_p A^2 \omega p \sin kp \quad (24.8)$$

The *energy velocity* is defined as in Eq. (20.11) and accordingly is

$$U_e = \frac{\Phi}{\bar{E}} = \frac{\sum_p \frac{1}{2} U''_p A^2 \omega p \sin kp}{\frac{m}{2a} A^2 \omega^2} = \frac{d}{m\omega} \sum_p U''_p p \sin kp \quad (24.9)$$

It is easy to verify that this is the same as the *group velocity*, for

$$U_g = \frac{\partial \nu}{\partial |a|} = d \frac{\partial \omega}{\partial k} = \frac{d}{2\omega} \frac{\partial \omega^2}{\partial k} \quad \text{since} \quad \begin{aligned} \omega &= 2\pi\nu \\ k &= 2\pi a d \end{aligned}$$

We have already shown [Eq. (10.1)] that

$$\omega^2 = \frac{2}{m} \sum_p U''_p (1 - \cos kp)$$

and hence

$$\frac{\partial \omega^2}{\partial k} = \frac{2}{m} \sum_p U''_p p \sin kp$$

and substitution yields the equation

$$U_g = \frac{d}{m\omega} \sum_p U''_p p \sin kp = U_e \quad (24.10)$$

In the previous section we defined the *characteristic impedance* by

$$f_n = -Z\dot{y}_n \quad (24.11)$$

where f_n is the force acting on particle n and \dot{y}_n its velocity. Here, however, we have more than one particle affected by the particles to the right of particle n , and thus the characteristic impedance cannot be defined by Eq. (24.11) since it is not the impedance that would be required to terminate the lattice at particle n in such a way that no reflection occurs. It would be necessary to combine the impedances due to the different L particles near the end of the line. A convenient way of doing this is offered by Eq. (23.2). According to this equation

$$\Phi = \frac{1}{2} Z A^2 \omega^2 \quad (24.12)$$

where Z is the characteristic impedance and always remains real. The combination of the impedances offered by the differ-

ent particles is implicit in the equation since the contribution of each particle to the total energy flow has been taken into consideration. We note that Eq. (24.12) is equivalent to

$$\Phi = \frac{1}{2} Z |\dot{y}_n|^2 \quad (24.13)$$

and we use Eq. (24.13) as the general definition of the characteristic impedance of a one-dimensional lattice. This definition introduces no inconsistencies, and once we have computed Φ , we need consider only the particle terminating the lattice in problems of finite lattices and the junction of lattices. Z is to be real and the cells so chosen that any imaginary part of Z that might arise is taken care of by the interactions of the particles in the lattice. This definition of the *characteristic impedance* enables one to state the *necessary* condition for *no reflection* at the end of the lattice: that the lattice be terminated on a system of impedances resulting in a total impedance equal to the characteristic impedance. This condition is *necessary but not sufficient*, and the general situation near the boundary is very similar to the one obtained in Secs. 8 and 10, where the problem of steady-state equilibrium was discussed.

For cases where the cells contain particles of various masses the problem is more complicated, but the same general methods are applicable. To obtain the average potential and kinetic energies, the average contributions of each particle in the cell to the energy are summed and divided by the length of the cell. The energy flow Φ is obtained by summing the contributions of all particles through a junction between cells. Once the flux Φ and the energy density \bar{E} are obtained, the energy velocity U_e is defined by Eq. (20.11) and found equal to the group velocity [Eq. (21.5)] for all structures exhibiting no absorption.

The curve $\nu(a)$ always has a horizontal tangent on the limits of the interval $a = \pm 1/2d$. This means $\partial\nu/\partial a = 0$ and zero group velocity and checks very well with the fact that these special waves behave practically like standing waves. This was observed, for instance, in the NaCl problem discussed in Chap. IV, for the waves corresponding to ω_1 and ω_2 , the limits of the passing bands.

The *characteristic impedance* becomes increasingly difficult to define and loses more and more of its practical significance. Different values would be found for Z according to the assump-

tions made about the distribution of Z between the particles of the last cell, and, furthermore, the Z value gives a necessary but not a sufficient condition for no reflection.

To conclude, let us emphasize the importance of quantities such as energy flow, energy density, energy velocity, and group velocity, in addition to the usual phase velocity. These definitions were introduced long ago by theoretical physicists. They can be extended from one to two or three dimensions. For electromagnetic waves, for instance, a most general definition of the energy flow leads to Poynting's vector.

The characteristic or surge impedance, familiar to electrical engineers, is very useful for one-dimensional structures with interactions between nearest neighbors only. This includes practically all problems of filters, lines, and cables for electrical communications. We have just found how delicate is the extension to one-dimensional structures with interactions at large distances. Despite many interesting attempts, the extension to two or three dimensions remains rather artificial.¹

¹ See SCHELKUNOFF, S. A., "Electromagnetic Waves," Chap. XII, Van Nostrand, New York, 1943.

CHAPTER VI

TWO-DIMENSIONAL LATTICES

25. Direct and Reciprocal Lattices in Two Dimensions

Lattices in two dimensions offer much the same sort of difficulties as those in three dimensions, but they are easier to discuss since drawings are simpler and clearer to understand. This is

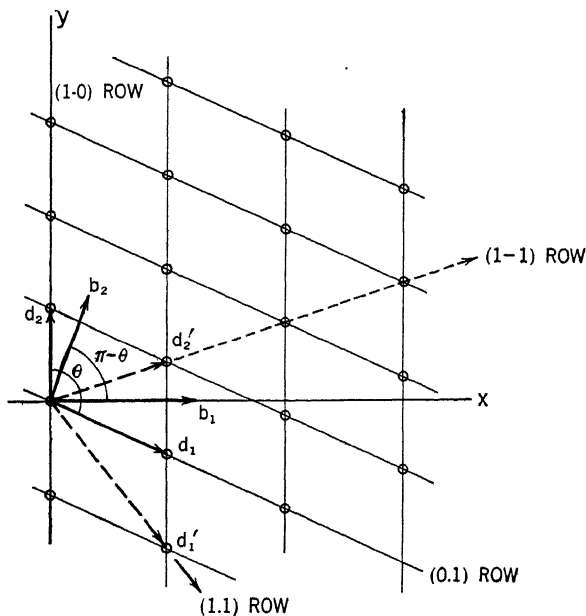


FIG. 25.1.

the reason why a whole chapter is devoted to the two-dimensional problem.

We shall start with a two-dimensional lattice composed of particles all having the same mass and spaced at equal distances from one another along two lines intersecting at an arbitrary angle θ . Later we shall find the generalization to more complicated lattices easy to make. This lattice is shown in Fig. 25.1.

The dots represent the masses. The distance between particles in direction \mathbf{d}_1 is not necessarily the same as in direction \mathbf{d}_2 .

We take \mathbf{d}_1 and \mathbf{d}_2 as basis vectors drawn from the particle chosen as the origin of the lattice. The vector coordinate of any point in the lattice is then given by

$$\mathbf{r}_{l_1 l_2} = l_1 \mathbf{d}_1 + l_2 \mathbf{d}_2 \quad (25.1)$$

where l_1 and l_2 are integers. The basis system is not, of course, unique. \mathbf{d}_1 and \mathbf{d}'_2 would serve just as well and, in fact, any two linearly independent vectors \mathbf{d}''_1 and \mathbf{d}''_2 given by

$$\left. \begin{aligned} \mathbf{d}''_1 &= m_1 \mathbf{d}_1 + n_1 \mathbf{d}_2 \\ \mathbf{d}''_2 &= m_2 \mathbf{d}_1 + n_2 \mathbf{d}_2 \end{aligned} \right\} \quad \frac{m_1}{m_2} \neq \frac{n_1}{n_2} \quad (25.2)$$

where m_1 , m_2 , n_1 , and n_2 are integers,¹ would give a satisfactory basis. The \mathbf{d}'' 's correspond to $n\mathbf{d}$ in the one-dimensional system, and there we used $n = 1$, i.e., the vector with the smallest absolute value greater than zero. Similarly, here we shall more or less arbitrarily designate \mathbf{d}_1 and \mathbf{d}_2 , the smallest pair of basis vectors greater than zero, as the basis vectors of the lattice.

When we assume two basis vectors \mathbf{d}_1 and \mathbf{d}_2 for a lattice, then the lattice is completely determined if we restrict ourselves to a single type of particle and require that the particles be equally spaced along the two independent directions. We refer to the lattice described by the vectors \mathbf{d}_1 and \mathbf{d}_2 as the *direct lattice*. For each direct lattice we may define a reciprocal lattice that is to have basis vectors \mathbf{b}_1 and \mathbf{b}_2 given by the equation

$$(\mathbf{b}_i \cdot \mathbf{d}_k) = \delta_{ik} \quad \left\{ \begin{array}{l} i = 1, 2 \\ k = 1, 2 \end{array} \right\} \quad (25.3)$$

where δ_{ik} is the Kronecker δ symbol, defined by

$$\delta_{ik} = \left\{ \begin{array}{ll} 1 & i = k \\ 0 & i \neq k \end{array} \right\} \quad (25.4)$$

The reasons for the term *reciprocal lattice* become apparent with a little calculation. For, if we take the origin of a pair of orthogonal axes x and y at the origin of the basis system, the vectors \mathbf{d}_1 and \mathbf{d}_2 may be written in terms of their Cartesian components as follows:

¹ The area of the new cell $\mathbf{d}''_1 \mathbf{d}''_2$ should equal that of $\mathbf{d}_1 \mathbf{d}_2$; otherwise the simple lattice is changed into a lattice with basis (see p. 128).

$$\left. \begin{aligned} \mathbf{d}_1 &= (d_{1x} \quad d_{1y}) \\ \mathbf{d}_2 &= (d_{2x} \quad d_{2y}) \end{aligned} \right\} \quad (25.5)$$

or, in terms of the matrix notation, the matrix

$$D = \begin{pmatrix} d_{1x} & d_{1y} \\ d_{2x} & d_{2y} \end{pmatrix} \quad (25.6)$$

represents the basis system. Similarly, the reciprocal basis system has the matrix

$$B = \begin{pmatrix} b_{1x} & b_{1y} \\ b_{2x} & b_{2y} \end{pmatrix} \quad (25.7)$$

The subscripts on the elements of the matrix D must be transposed for the matrix B since, if \mathbf{d}_1 and \mathbf{d}_2 are thought of as *row* vectors, \mathbf{b}_1 and \mathbf{b}_2 must be *column* vectors because the latter are defined by taking a scalar product with the former [Eq. (25.3)]. It is readily verified that D and B are *reciprocal matrices*. Let us form the matrix product.

$$\begin{aligned} D \cdot B &= \begin{pmatrix} d_{1x} & d_{1y} \\ d_{2x} & d_{2y} \end{pmatrix} \begin{pmatrix} b_{1x} & b_{1y} \\ b_{2x} & b_{2y} \end{pmatrix} \\ &= \begin{pmatrix} d_{1x}b_{1x} + d_{1y}b_{1y} & d_{1x}b_{2x} + d_{1y}b_{2y} \\ d_{2x}b_{1x} + d_{2y}b_{1y} & d_{2x}b_{2x} + d_{2y}b_{2y} \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{d}_1 \cdot \mathbf{b}_1) & (\mathbf{d}_1 \cdot \mathbf{b}_2) \\ (\mathbf{d}_2 \cdot \mathbf{b}_1) & (\mathbf{d}_2 \cdot \mathbf{b}_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \delta \end{aligned} \quad (25.8)$$

where δ is the unit matrix. From this it follows that

$$B = D^{-1} \quad (25.9)$$

From Eq. (25.3) we see that

$$\left. \begin{aligned} \mathbf{b}_1 &\text{ is perpendicular to } \mathbf{d}_2 \\ \mathbf{b}_2 &\text{ is perpendicular to } \mathbf{d}_1 \end{aligned} \right\} \quad (25.10)$$

and therefore

$$\left. \begin{aligned} (\mathbf{b}_1 \cdot \mathbf{d}_1) &= 1 = |\mathbf{b}_1||\mathbf{d}_1| \cos \left(\frac{\pi}{2} - \theta \right) = |\mathbf{b}_1||\mathbf{d}_1| \sin \theta \\ (\mathbf{b}_2 \cdot \mathbf{d}_2) &= 1 = |\mathbf{b}_2||\mathbf{d}_2| \cos \left(\frac{\pi}{2} - \theta \right) = |\mathbf{b}_2||\mathbf{d}_2| \sin \theta \end{aligned} \right\} \quad (25.11)$$

where θ is the angle between \mathbf{d}_1 and \mathbf{d}_2 . This is easily seen by inspection of Fig. 25.1. Now the area of the elementary cell in

the direct lattice (*i.e.*, the parallelogram with \mathbf{d}_1 and \mathbf{d}_2 for two of its sides) is given by

$$S_d = |\mathbf{d}_1 \times \mathbf{d}_2| = |\mathbf{d}_1||\mathbf{d}_2| \sin \theta \quad (25.12)$$

from elementary vector analysis, while that of the elementary cell in the reciprocal lattice is

$$S_b = |\mathbf{b}_1||\mathbf{b}_2| \sin \theta \quad (25.13)$$

The product of these areas is

$$\begin{aligned} S_d \cdot S_b &= |\mathbf{d}_1||\mathbf{d}_2| \sin \theta |\mathbf{b}_1||\mathbf{b}_2| \sin \theta \\ &= |\mathbf{d}_1||\mathbf{d}_2| \sin \theta \frac{\sin \theta}{|\mathbf{d}_1| \sin \theta |\mathbf{d}_2| \sin \theta} = 1 \end{aligned} \quad (25.14)$$

from Eq. (25.11). In other words, the *areas* of the direct and reciprocal cells are reciprocals.

In the one-dimensional lattice the length of the cell in the direct lattice was d . The length of the cell in the frequency vs. wave-number space was $1/d$, and this was, therefore, the reciprocal cell of the lattice. Thus the direct lattice gave the periodicity of the medium, and the reciprocal lattice gave that of the frequency of the waves propagating through the medium. Similar results will be obtained for two dimensions.

For readers accustomed to the definitions of tensor analysis, the following comment may be added. The \mathbf{d}_1 and \mathbf{d}_2 basis vectors of the direct lattice play essentially the role of the covariant unit vectors, while \mathbf{b}_1 and \mathbf{b}_2 represent the contravariant unit vectors in an oblique coordinate system.¹

As a matter of fact, many discussions are simplified if the \mathbf{d}_1 and \mathbf{d}_2 vectors are used as unit vectors defining an oblique axis system, and any arbitrary vector \mathbf{r} is given by its ξ_1 and ξ_2 components along the \mathbf{d} vectors.

$$\mathbf{r} = \xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2 \quad (25.15)$$

The oblique $(\mathbf{d}_1, \mathbf{d}_2)$ cell in the xy space is thus reduced to a square cell in the ξ space, since

$$\begin{array}{lll} \text{vector } \mathbf{d}_1 \text{ means} & \xi_1 = 1, & \xi_2 = 0 \\ \text{vector } \mathbf{d}_2 \text{ means} & \xi_1 = 0, & \xi_2 = 1 \end{array}$$

¹ BRILLOUIN, L., "Les Tenseurs en mécanique et en élasticité," pp. 27-30, 97, 101, 105, Masson, Paris, 1938.

STRATTON, J. A., "Electromagnetic Theory," p. 39, McGraw-Hill, New York, 1942.

Furthermore, according to the defining equation (25.3)

$$\xi_1 = (\mathbf{r} \cdot \mathbf{b}_1), \quad \xi_2 = (\mathbf{r} \cdot \mathbf{b}_2) \quad (25.16)$$

A straight line in the plane (this line will become a plane in three dimensions) is represented by a linear relation.

$$\left. \begin{aligned} (\mathbf{a} \cdot \mathbf{r}) &= a_1x + a_2y = c \\ \text{or } \alpha_1\xi_1 + \alpha_2\xi_2 &= c \end{aligned} \right\} \quad \begin{aligned} \alpha_1 &= (\mathbf{a} \cdot \mathbf{d}_1) \\ \alpha_2 &= (\mathbf{a} \cdot \mathbf{d}_2) \end{aligned} \quad (25.17)$$

as is easily seen by direct substitution. Conversely,

$$\mathbf{a} = \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 \quad (25.18)$$

In these equations \mathbf{a} obviously represents a vector orthogonal to the straight line (25.17), and $c/|\mathbf{a}|$ is the distance δ of the line from the origin.

A lattice point is one with integral coordinates l_1 and l_2 .

$$\xi_1 = l_1, \quad \xi_2 = l_2, \quad \mathbf{r} = l_1\mathbf{d}_1 + l_2\mathbf{d}_2 \quad (25.19)$$

and a vector \mathbf{h} in the reciprocal lattice is

$$\mathbf{h} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 \quad h_1 \text{ and } h_2 \text{ integers} \quad (25.20)$$

If such an \mathbf{h} vector is taken as vector \mathbf{a} in Eq. (25.17), it defines (for different values of c) a set of parallel lines, some of which go through an infinite number of points of the direct lattice. One of these lines is

$$(\mathbf{h} \cdot \mathbf{r}) = h_1(\mathbf{b}_1 \cdot \mathbf{r}) + h_2(\mathbf{b}_2 \cdot \mathbf{r}) = h_1\xi_1 + h_2\xi_2 = c \quad (25.21)$$

For $c = 0$ the line passes through the origin and through all lattice points for which

$$h_1l_1 + h_2l_2 = 0$$

such as

$$l_1 = h_2 \quad \text{and} \quad l_2 = -h_1$$

Other lattice rows will correspond to different c values. Now we ask the following question: What is the distance from the origin of the lattice row in this set nearest to the origin? This is the same as asking: What is the smallest nonzero value of $|c|$? Since h_1 , ξ_1 , h_2 , and ξ_2 are all integers (positive or negative), c is also an integer for a lattice row, and the smallest nonzero value of $|c|$ is, provided h_1 h_2 have no common factor and \mathbf{h} is the smallest possible vector.

$$|c| = 1$$

This means that the distance between each of the lattice rows in the set (h_1, h_2) is

$$\delta = \frac{|c|}{|\mathbf{h}|} = \frac{1}{|\mathbf{h}|} \quad (25.22)$$

Thus we have the following statement: A point (h_1, h_2) in the reciprocal lattice defines a set of lattice rows in the direct lattice. These straight rows are perpendicular to the vector \mathbf{h} and are spaced at a distance of $1/|\mathbf{h}|$ from one another. This is known as the *Bravais* notation for crystal planes. In Fig. 25.1, the $(1,0)$ rows are the vertical lines, and the $(0,1)$ set is parallel to \mathbf{d}_1 . The vector \mathbf{d}'_1 is in a $(1,1)$ row, while \mathbf{d}'_2 is in $(1, -1)$.

According to the physical properties to be discussed, sometimes the direct and at other times the reciprocal lattice will yield the better description of the periodic structure.

26. Doubly and Triply Periodic Functions

We shall have use for doubly periodic functions only in this chapter. Since, however, triply periodic functions will arise in the theory of three-dimensional lattices, we shall treat the two together. By a doubly or triply periodic function we mean a function of two or three independent variables, periodic in each of its variables. The mathematical theory of such functions would be considerably simplified if we could split an arbitrary function, say $D(x, y)$, that is periodic in x and y into a product of two functions $F_1(x)$ and $F_2(y)$ periodic in x and y , respectively. This can be done, however, only in very special cases. Suppose, for instance, that

$$F(x, y) = \begin{cases} 1 & x, y \text{ both integers,} \\ 0 & \text{otherwise} \end{cases} \quad (26.1)$$

Figure 26.1 shows $F(x, y)$ plotted in the xy plane. The dots represent points at which $F(x, y)$ is not zero. Then

$$\begin{aligned} F_1(x) &= \begin{cases} 1 & x \text{ integer} \\ 0 & \text{otherwise} \end{cases} \\ F_2(y) &= \begin{cases} 1 & y \text{ integer} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (26.2)$$

are two functions of one variable each whose product is $F(x, y)$. Evidently $F_1(x)$ and $F_2(y)$ are both periodic functions. Consider, however, the function

$$F(x,y) = \begin{cases} 1 & x \text{ or } y \text{ integer} \\ 0 & \text{neither } x \text{ nor } y \text{ integer} \end{cases} \quad (26.3)$$

Figure 26.2 shows the function $F(x,y)$ plotted on the xy plane. The horizontal and vertical lines represent the points at which $F(x,y)$ is not zero. Let us assume that

$$F(x,y) = F_1(x)F_2(y) \quad (26.4)$$

holds for all x and y . Then, if x_1 and y_1 are neither integers,

$$F_1(x_1)F_2(y_1) = 0 \quad (26.5)$$

so that

$$F_1(x_1) = 0 \quad \text{or} \quad F_2(y_1) = 0 \quad (26.6)$$

Let us suppose $F_2(y_1) = 0$; then

$$F_1(n)F_2(y_1) = 0 \quad n \text{ an integer} \quad (26.7)$$

contradicts the hypothesis that $F(x,y)$ defined by Eq. (26.3) has the form (26.4). Starting from $F_1(x_1)$ would lead to a similar

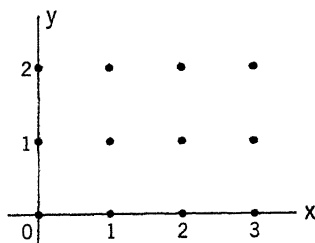


FIG. 26.1.

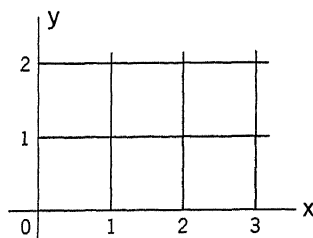


FIG. 26.2.

conclusion; hence the decomposition is impossible. Many authors have also tried a sum.

$$F(x,y) = F_1(x) + F_2(y) \quad (26.8)$$

An example of this would be given by taking F_1 and F_2 as in Eq. (26.2).

$$F(x,y) = \begin{cases} 2 & x \text{ and } y \text{ integers} \\ 1 & x \text{ or } y \text{ integer} \\ 0 & \text{elsewhere} \end{cases}$$

Such an example is obviously a very special case of little practical use.

We shall, however, find it possible to expand doubly or triply periodic functions in double or triple Fourier series, and this will be done by means of the reciprocal lattice defined in Sec. 25.

A periodic function $F(x, y)$ in the direct lattice has the same value at points

$$\mathbf{r} = (x, y) \quad \text{and} \quad \mathbf{r}' = \mathbf{r} + l_1 \mathbf{d}_1 + l_2 \mathbf{d}_2 \quad (26.9)$$

Changing to coordinates ξ_1 and ξ_2 as in Eq. (25.15), we obtain a function $f(\xi_1, \xi_2)$ with period 1 in both ξ_1 and ξ_2 . This periodic function can be expanded in a double Fourier series of imaginary exponentials.

$$F(x, y) = f(\xi_1, \xi_2) = \sum_{h_1 h_2} C_{h_1 h_2} e^{2\pi i (h_1 \xi_1 + h_2 \xi_2)} \quad (26.10)$$

where h_1 and h_2 are integers, and the coefficients of this series are given by

$$\begin{aligned} \int_0^1 \int_0^1 f(\xi_1, \xi_2) e^{-2\pi i (h'_1 \xi_1 + h'_2 \xi_2)} d\xi_1 d\xi_2 \\ = \sum_{h_1 h_2} C_{h_1 h_2} \int_0^1 \int_0^1 e^{2\pi i [(h_1 - h'_1) \xi_1 + (h_2 - h'_2) \xi_2]} d\xi_1 d\xi_2 \end{aligned}$$

The integral is zero whenever $h_1 \neq h'_1$ or $h_2 \neq h'_2$ and the only remaining term is the one for which $h_1 = h'_1$ and $h_2 = h'_2$.

$$C_{h_1 h_2} = \int_0^1 \int_0^1 f(\xi_1, \xi_2) e^{-2\pi i (h_1 \xi_1 + h_2 \xi_2)} d\xi_1 d\xi_2 \quad (26.11)$$

The $C_{h_1 h_2}$ coefficient is generally complex and includes the amplitude and phase angles. Returning to the original $F(x, y)$ function and making use of Eq. (25.21), we obtain

$$F(x, y) = \sum_{h_1 h_2} C_{h_1 h_2} e^{2\pi i (\mathbf{h} \cdot \mathbf{r})} = \sum_{h_1 h_2} C_{h_1 h_2} e^{2\pi i (h_1 (b_1 \cdot \mathbf{r}) + h_2 (b_2 \cdot \mathbf{r}))} \quad (26.12)$$

$$\mathbf{h} = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2$$

where \mathbf{h} defines one of the vectors of the reciprocal lattice. The periodic character of this expansion can be checked directly from Eqs. (25.20) and (25.3).

$$\begin{aligned} (\mathbf{h} \cdot \mathbf{r}') &= [\mathbf{h} \cdot (\mathbf{r} + l_1 \mathbf{d}_1 + l_2 \mathbf{d}_2)] \\ &= (\mathbf{h} \cdot \mathbf{r}) + h_1 l_1 (\mathbf{b}_1 \cdot \mathbf{d}_1) + h_2 l_2 (\mathbf{b}_2 \cdot \mathbf{d}_2) = (\mathbf{h} \cdot \mathbf{r}) + h_1 l_1 + h_2 l_2 \end{aligned}$$

The imaginary exponential in Eq. (26.12) obviously has the same value at $(\mathbf{h} \cdot \mathbf{r})$ as at $[(\mathbf{h} \cdot \mathbf{r}) + \text{integer}]$.

In the transcription of Eq. (26.11) we must be cautious, since ξ_1 and ξ_2 correspond to oblique coordinates, and the length of a vector \mathbf{r} is

$$|\mathbf{r}|^2 = x^2 + y^2 = |\mathbf{d}_1|^2 \xi_1^2 + 2(\mathbf{d}_1 \cdot \mathbf{d}_2) \xi_1 \xi_2 + |\mathbf{d}_2|^2 \xi_2^2 \quad (26.13)$$

The area S_d of the \mathbf{d} cell is transformed into an area 1 in the ξ system; hence

$$S_d = |\mathbf{d}_1| |\mathbf{d}_2| \sin \theta, \quad dx dy = S_d d\xi_1 d\xi_2 \quad (26.14)$$

and

$$C_{h_1 h_2} = \frac{1}{S_d} \int_0^{d_1} \int_0^{d_2} F(x, y) e^{-2\pi i (\mathbf{h} \cdot \mathbf{r})} dx dy \quad (26.15)$$

To give a physical interpretation to expansion (26.12), we may say that the periodic function is decomposed into plane waves, corresponding to each of the lattice rows (lattice planes in three dimensions) defined by the points \mathbf{h} in the reciprocal lattice. A periodic function $F(x, y)$ that can be represented by a product $F_1(x)F_2(y)$ offers the very special property that

$$C_{h_1 h_2} = C_{h_1} C_{h_2} \quad \text{with} \quad C_{h_1} = \int_0^1 f_1(\xi_1) e^{-2\pi i h_1 \xi_1} d\xi_1 \quad (26.16)$$

and a similar equation for C_{h_2} . This, obviously, cannot be general, but it retains the whole set of coefficients C . The assumption in Eq. (26.8) of a sum $F_1(x) + F_2(y)$ is much more restricting since it knocks out all the coefficients except $C_{h_1 0}$ and $C_{0 h_2}$, which means that all oblique atomic rows are ruled out.

To emphasize the importance of these definitions, let us state that X-ray reflections from a crystal yield directly the values of $|C_{h_1 h_2}|$ in the expansion of electronic density inside the crystal. Only the absolute value of the coefficients is obtained from the experimental data, but symmetry considerations often enable one to guess the phase angles and to reconstruct the whole periodic function representing the average density of electrons throughout the crystal lattice.

27. Zones in a Two-dimensional Lattice

In the discussion of the one-dimensional case, we found that the frequency was a periodic function of the wave number, and hence for a given frequency there was ambiguity in the wave length and the direction of propagation. We chose an interval containing one period of the frequency and taken symmetrically

about the origin, and then we restricted the wave number to values in this interval. In the language of this section, we should call this interval the first *zone*. The second zone in one dimension would consist of two intervals containing half a period each, one on each side of the first zone, etc., for higher order zones. The zones for one dimension are shown in Fig. 27.1.

We now proceed to find the analogues of these zones in the two-dimensional case. The zones will be regions in the reciprocal lattice, since this is the lattice describing the periodicity of

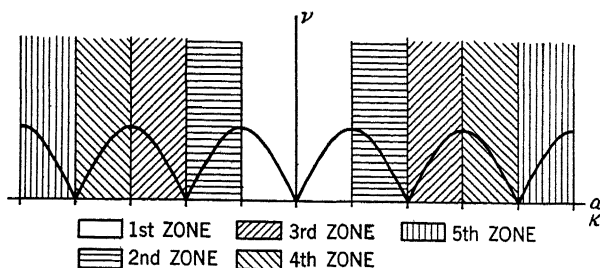


FIG. 27.1.—Zones in one dimension.

frequency as a function of wave number. Let us consider a plane wave propagating through the two-dimensional medium. It will have the form

$$\psi = A e^{i(\omega t - 2\pi a_1 x - 2\pi a_2 y)} \quad (27.1)$$

or if we let \mathbf{a} be a vector with components a_1 and a_2 and \mathbf{r} a vector with components x and y ,

$$\psi = A e^{i[\omega t - 2\pi(\mathbf{a} \cdot \mathbf{r})]} \quad \text{and} \quad |\mathbf{a}|^2 = a_1^2 + a_2^2 = \frac{1}{\lambda^2} \quad (27.2)$$

\mathbf{a} will be a vector in the direction of propagation, and its magnitude will be the reciprocal of the wave length λ . In a discontinuous medium, ψ is defined only at points \mathbf{r} at which particles are located; *i.e.*, at the points $\mathbf{r}_{l_1 l_2}$ of the direct lattice. Thus

$$2\pi(\mathbf{a} \cdot \mathbf{r}_{l_1 l_2}) = 2\pi(\mathbf{a} \cdot l_1 \mathbf{d}_1) + 2\pi(\mathbf{a} \cdot l_2 \mathbf{d}_2) = l_1 k_1 + l_2 k_2 \quad (27.3)$$

where we have set

$$k_1 = 2\pi(\mathbf{a} \cdot \mathbf{d}_1) \quad \text{and} \quad k_2 = 2\pi(\mathbf{a} \cdot \mathbf{d}_2) \quad (27.4)$$

Accordingly, Eq. (27.2) may be written

$$\psi = A e^{i(\omega t - k_1 l_1 - k_2 l_2)} \quad (27.5)$$

Now, as in the one-dimensional case, we may replace k_1 and k_2 by k'_1 and k'_2 where

$$k'_i = k_i + 2\pi m_i \quad m_i \text{ integers} \quad (27.6)$$

without changing either the motion of the particles or the frequency. This ambiguity in the value of k_1 and k_2 is the analogue of that in k in the one-dimensional case (Sec. 4). The values of k'_1 and k'_2 in Eq. (27.6) correspond to the following value for \mathbf{a}' :

$$\mathbf{a}' = \mathbf{a} + m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2 \quad (27.7)$$

for

$$\left. \begin{aligned} k'_1 &= 2\pi(\mathbf{a}' \cdot \mathbf{d}_1) = 2\pi(\mathbf{a} \cdot \mathbf{d}_1) + 2\pi m_1(\mathbf{b}_1 \cdot \mathbf{d}_1) \\ &\quad + 2\pi m_2(\mathbf{b}_2 \cdot \mathbf{d}_1) = k_1 + 2\pi m_1 \\ k'_2 &= 2\pi(\mathbf{a}' \cdot \mathbf{d}_2) = 2\pi(\mathbf{a} \cdot \mathbf{d}_2) + 2\pi m_1(\mathbf{b}_1 \cdot \mathbf{d}_2) \\ &\quad + 2\pi m_2(\mathbf{b}_2 \cdot \mathbf{d}_2) = k_2 + 2\pi m_2 \end{aligned} \right\} \quad (27.8)$$

The direction of propagation is given by \mathbf{a}' , and this, as well as the magnitude of the wave length, will depend on m_1 and m_2 in Eq. (27.7). Various \mathbf{a}' vectors corresponding to a given \mathbf{a}

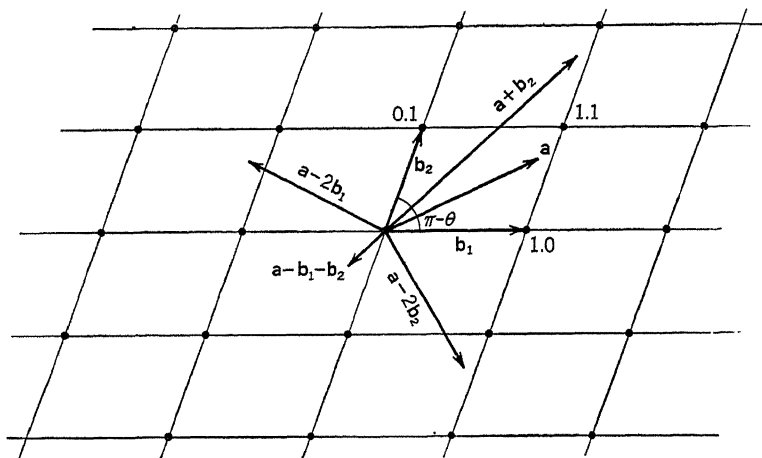


FIG. 27.2.

are shown drawn in the reciprocal lattice in Fig. 27.2. This discussion shows that the frequency ν of the wave, in a two-dimensional lattice, is a periodic function of the wave vector \mathbf{a} in the reciprocal lattice with basis vectors \mathbf{b}_1 and \mathbf{b}_2 .

We must now formulate a rule for choosing the area to which \mathbf{a} is to be confined. We might start at the origin of the basis

system and confine all vectors \mathbf{a} to the first elementary cell of the reciprocal lattice—*i.e.*, to the parallelogram bounded by the points (0,0), (0,1), (1,0), and (1,1) in Fig. 27.2. There are two disadvantages to this. In the first place, this method singles out the directions contained in the angle $\pi - \theta$ as preferred directions of propagation for the waves. Further, it does not require the use of the longest possible wave length for a given disturbance and thus is inconsistent with the conventions set up for the one-dimensional case.

To do away with these objections, we try to construct a zone that will be analogous to the first zone in the one-dimensional

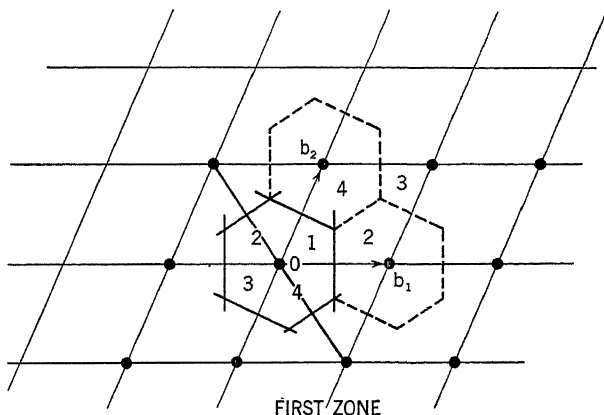


FIG. 27.3.

case. This means, first, that we must place the origin in the center of the zone. The remainder of the construction is accomplished by drawing perpendicular bisectors of the lines joining the origin to each of the other points in the reciprocal lattice. The smallest closed polygon formed by these perpendicular bisectors is taken as the first zone. It is independent of the basis system chosen and allows propagation in all directions. Furthermore, it requires the *longest wave length* describing a given disturbance to be used, as is easily seen by inspection since a complete period for each direction of propagation is included in the zone. There is still an ambiguity on the boundaries of the zone just as in the one-dimensional case. The construction for the first zone is shown in Fig. 27.3. The first zone has the same area as the first elementary cell of the reciprocal lattice, as

may be seen by comparing elements of the two areas containing the same numbers. These elements are easily seen to be congruent from plane geometry.

The construction of the second zone and higher order zones is more complicated. They must all be bounded by perpendicular bisectors of lines joining the origin with other lattice points, and there can be no perpendicular bisectors passing through the interior of a zone. The significance of this will be seen later.

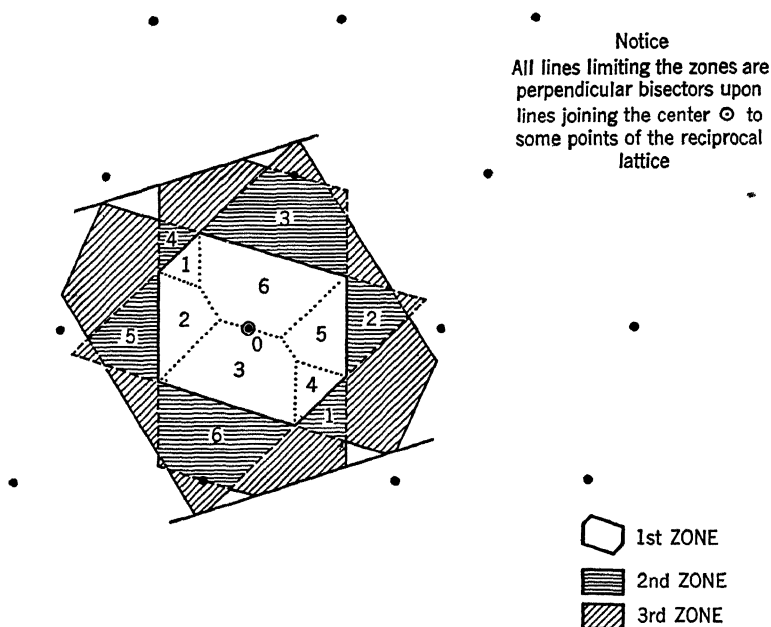


FIG. 27.4.

If a wave propagates through a continuous medium with small periodic variations in such a manner that its wave vector measured from the origin terminates on a perpendicular bisector of a line joining the origin with a lattice point, a discontinuity occurs in the ν vs. $|a|$ curve. The object of introducing zones is to eliminate discontinuities in the ν vs. $|a|$ curve except at the boundaries. This is exactly analogous to the one-dimensional case.

To construct the second zone, we draw the second smallest closed figure about the origin and bounded by perpendicular

bisectors. The second zone is the area enclosed between the boundary of the first zone and the boundary of this second figure. The construction is shown in Fig. 27.4. Similarly, one may construct a third zone that is the area enclosed between the boundary of the second zone and the third smallest closed figure bounded by perpendicular bisectors. This is also shown in Fig. 27.4. Higher order zones are constructed in just the same manner. It should be noted that the $(n + 1)$ st zone consists of figures having at least one side in common with one side of the n th zone and all vertices in common with the $(n - 1)$ st zone.

Each zone has the same area as the elementary cell in the reciprocal lattice. This is shown by taking sections of the zone under consideration and noting their position relative to some lattice point. Now the wave vectors terminating in the corresponding section of any other cell will give the same value for ψ , since the only change in ψ is the addition of $2m\pi$ in the exponent of the exponential. Therefore, we may consider the two sections equivalent. The matching of sections in the elementary cell with sections in the first and second zones is shown in Fig. 27.4. In all these cases the area of each of the zones is equal to the area of the elementary cell in the reciprocal lattice, and this is true for all cases that have been worked out. However, a general proof that this will always be the case has never been given.

28. Propagation of Waves in a Continuous Two-dimensional Medium with a Periodic Perturbation

We are considering waves propagating in a two-dimensional continuum with a nonuniformity in the structure of the medium that is periodic in each of the two independent directions. This problem was sketched in Sec. 17 for the one-dimensional case and will be fully discussed now. We can define a direct lattice with basis vectors pointing in the two directions \mathbf{d}_1 and \mathbf{d}_2 and having magnitudes equal to the periods of the structure in these two directions. Each point in the first elementary cell may be defined by a vector of the form [Eq. (25.15)]

$$\mathbf{r} = \xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2 \quad \begin{array}{l} 0 < \xi_1 < 1 \\ 0 < \xi_2 < 1 \end{array}$$

The vectors from the origin of the lattice for each point in the lattice are obtained by adding the vector for the corresponding

point in the first cell to the vector of the origin of the cell in which the point lies.

$$\mathbf{r}' = (\xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2) + m_1 \mathbf{d}_1 + m_2 \mathbf{d}_2$$

A reciprocal lattice with basis vectors \mathbf{b}_1 and \mathbf{b}_2 defined by

$$(\mathbf{b}_i \cdot \mathbf{d}_k) = \delta_{ik}$$

as in Sec. 25 may be constructed. The vectors in this lattice are the propagation vectors of a wave propagating in the direct lattice, and the frequency of the wave is a function of the propagation vectors.

The general wave equation for a two-dimensional continuum with periodic nonuniformities is

$$\nabla^2 \psi - \frac{1}{V^2} \ddot{\psi} = 0 \quad (28.1)$$

where ∇^2 is the two-dimensional Laplacian operator, ψ is the wave function, and V is the phase velocity of the wave, assumed to be a periodic function with periods \mathbf{d}_1 and \mathbf{d}_2 . ψ is a function of (x, y) and of the time t . We assume that the time-dependent part of ψ is separable from the space-dependent part, *i.e.*,

$$\psi = u(x, y) e^{i\omega t} \quad (28.2)$$

from which we obtain a differential equation for u .

$$\nabla^2 u + \frac{\omega^2}{V^2} u = 0 \quad (28.3)$$

Now the wave equation for a two-dimensional homogeneous isotropic continuum is (after the time part is eliminated)

$$\nabla^2 u_0 + \frac{\omega_0^2}{V_0^2} u_0 = 0 \quad (28.4)$$

where V_0 is a constant depending on the constants of the medium. A solution of Eq. (28.4) is

$$u_0 = A e^{-2\pi i(a \cdot r)} \quad (28.5)$$

where

$$4\pi^2 |a|^2 = \frac{\omega_0^2}{V_0^2}, \quad |a|^2 = \frac{\nu_0^2}{V_0^2} \quad (28.6)$$

Equation (28.5) represents a plane wave. We may think of the

wave as an ordinary sound wave or as an electromagnetic wave. Both of these have constant velocities in a homogeneous medium.

We solve the problem of propagation of waves through a continuum with periodic structure by perturbation methods. Thus we set

$$\left. \begin{aligned} u &= u_0 + \epsilon u_1 \\ \omega^2 &= \omega_0^2 + \epsilon k_1 \\ \frac{1}{V^2} &= \frac{1}{V_0^2} + \epsilon f \end{aligned} \right\} \quad (28.7)$$

We take \mathbf{a} , the propagation vector, as fixed both in magnitude and in direction. u , $1/V^2$, and ω^2 are all functions of \mathbf{a} in the reciprocal and \mathbf{r} in the direct lattice. ϵ is a constant small enough so that any terms in ϵ^2 , ϵ^3 , . . . that we may encounter may be neglected in comparison with terms in ϵ , at least to a first approximation. The function f is periodic, and its average is zero in order to ensure

$$\frac{1}{V_0^2} = \overline{\left(\frac{1}{V^2} \right)}$$

We substitute Eq. (28.7) in the general wave equation (28.3).

$$\begin{aligned} \nabla^2 u + \frac{\omega^2}{V^2} u = 0 &= \nabla^2 u_0 + \frac{\omega_0^2}{V_0^2} u_0 \\ &+ \epsilon \left[\nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 + \left(\frac{k_1}{V_0^2} + \omega_0^2 f \right) u_0 \right] \end{aligned} \quad (28.8)$$

if we neglect terms in ϵ^2 and higher powers of ϵ . The zero-order approximation is obtained by neglecting the term in ϵ . This is just the wave equation for the continuum without variations, and the solution has already been given. The first-order correction to the zero-order approximation is given by equating the term in ϵ to zero.

$$\nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 = - \left(\frac{k_1}{V_0^2} + \omega_0^2 f \right) u_0 \quad (28.9)$$

This is an inhomogeneous differential equation as it stands. The homogeneous differential equation in u_1 is identical with the zero-order approximation and has the solutions

$$u_1 = B e^{-2\pi i(a'' \cdot r)} \quad (28.10)$$

where

$$4\pi^2|\mathbf{a}''|^2 = \frac{\omega_0^2}{V_0^2} = 4\pi^2|\mathbf{a}|^2, \quad |\mathbf{a}''|^2 = |\mathbf{a}|^2 \quad (28.11)$$

To obtain the solution of the inhomogeneous equation we expand the periodic function f in a double Fourier series as in Sec. 26.

$$f = \sum C_{m_1 m_2} e^{2\pi i(m_1 \xi_1 + m_2 \xi_2)} \quad (28.12)$$

where the $C_{m_1 m_2}$ are constants (in general complex to take care of phase factors) that may be evaluated from Fourier's theorem and $C_{00} = 0$. ξ_1 and ξ_2 are the coordinates of the point in the direct lattice at which f is to be evaluated with respect to the origin of the cell in which it lies, and m_1 and m_2 are integers. We may write f as a function of the vectors in the reciprocal lattice as in Eq. (26.12) by noting that

$$\left. \begin{aligned} (\mathbf{b}_1 \cdot \mathbf{r}) &= (\mathbf{b}_1 \cdot \xi_1 \mathbf{d}_1) + (\mathbf{b}_1 \cdot \xi_2 \mathbf{d}_2) = \xi_1 \\ (\mathbf{b}_2 \cdot \mathbf{r}) &= (\mathbf{b}_2 \cdot \xi_1 \mathbf{d}_1) + (\mathbf{b}_2 \cdot \xi_2 \mathbf{d}_2) = \xi_2 \end{aligned} \right\} \quad (28.13)$$

and, therefore,

$$f = \sum C_{m_1 m_2} e^{2\pi i[m_1(\mathbf{b}_1 \cdot \mathbf{r}) + m_2(\mathbf{b}_2 \cdot \mathbf{r})]} \quad (28.14)$$

Substituting Eq. (28.14) into Eq. (28.9), substituting the solution [Eq. (28.5)] of Eq. (28.4) for u_0 , and introducing the abbreviation

$$\mathbf{a}'_{m_1 m_2} = \mathbf{a} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2 \quad (28.15)$$

yields

$$\begin{aligned} \nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 &= -A \left(\frac{k_1}{V_0^2} e^{-2\pi i(\mathbf{a} \cdot \mathbf{r})} \right. \\ &\quad \left. + \omega_0^2 \sum C_{m_1 m_2} e^{-2\pi i(\mathbf{a}'_{m_1 m_2} \cdot \mathbf{r})} \right) = R(\mathbf{r}) \end{aligned} \quad (28.16)$$

This is an equation with right-hand term of the type

$$\nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 = R(\mathbf{r})$$

As is well known, such an equation possesses a *finite solution only* if the *right-hand term is orthogonal to all solutions of the homogeneous equation*, by which we understand the condition

$$\int_{\text{all space}} u_1^* R(\mathbf{r}) d\tau = 0$$

where u_1^* is the complex conjugate of u_1 in Eq. (28.10). In our problem this means that first of all we must write

$$\int \int_{\text{all space}} R(\mathbf{r}) e^{2\pi i(\mathbf{a}'' \cdot \mathbf{r})} d\mathbf{r} = 0, \quad |\mathbf{a}''| = |\mathbf{a}| \quad (28.17)$$

where \mathbf{a}'' is any vector of length $|\mathbf{a}|$ according to Eq. (28.11). Once this condition is satisfied, the general solution of Eq. (28.16) has the form

$$u_1 = \sum B_{m_1 m_2} e^{-2\pi i(\mathbf{a}'_{m_1 m_2} \cdot \mathbf{r})} \quad (28.18)$$

and we shall be able to evaluate the $B_{m_1 m_2}$ coefficients of this expansion in terms of known quantities. The general character of the solution will perhaps be better understood if it is written in a slightly different way.

$$u_1 = e^{-2\pi i(\mathbf{a} \cdot \mathbf{r})} F(\mathbf{r}); \quad F(\mathbf{r}) = \sum B_{m_1 m_2} e^{2\pi i[m_1(\mathbf{b}_1 \cdot \mathbf{r}) + m_2(\mathbf{b}_2 \cdot \mathbf{r})]}$$

This means that the solution ψ is a plane wave.

$$\psi = e^{i\omega t}(u_0 + \epsilon u_1) = e^{2\pi i[\nu t - (\mathbf{a} \cdot \mathbf{r})]} A(\mathbf{r}); \quad A(\mathbf{r}) = A + \epsilon F(\mathbf{r})$$

with an amplitude $A(\mathbf{r})$ that is a periodic function in the direct lattice.

In the discussion of condition (28.17) two cases must be distinguished:

1. Among all vectors \mathbf{a}'' having the same magnitude as vector \mathbf{a} , \mathbf{a} must be considered separately, but there is no other vector \mathbf{a}'' coinciding with any of the $\mathbf{a}'_{m_1 m_2}$ vectors.

$$\mathbf{a}'' \neq \text{all } \mathbf{a}'_{m_1 m_2}$$

2. An exact (or approximate) coincidence can be found for a certain vector \mathbf{a}'' and a corresponding $\mathbf{a}'_{m_1 m_2}$.

$$\mathbf{a}'' \approx \mathbf{a}'_{m_1 m_2} \quad \text{and} \quad \mathbf{a}'_{m_1 m_2} = \mathbf{a} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2$$

which implies the condition

$$|\mathbf{a}'_{m_1 m_2}| \approx |\mathbf{a}|$$

If this happens, both \mathbf{a} and this special \mathbf{a}'' must be singled out in the discussion of condition (28.17).

Case 1.—In this case we first write the necessary condition (28.17) for vector \mathbf{a} , and we obtain a relation that determines the value of the unknown coefficient k_1 . The integral in Eq. (28.17)

becomes

$$\int e^{2\pi i(a \cdot r)} \left(\frac{k_1}{V_0^2} e^{-2\pi i(a \cdot r)} + \omega_0^2 \sum C_{m_1 m_2} e^{-2\pi i(a' \cdot r)} \right) d\tau = \frac{k_1}{V_0^2} \int d\tau = 0$$

or

$$k_1 = 0 \quad (28.19)$$

since

$$\mathbf{a} - \mathbf{a}' = m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2$$

is zero only for $m_1 = m_2 = 0$ and $C_{00} = 0$, and hence all the exponential terms become zero on integration. Equation (28.19) means that the average perturbation on the unperturbed wave is zero for this case. For all other vectors \mathbf{a}' we have $\mathbf{a}' \neq \mathbf{a}$ and $|\mathbf{a}'| \neq |\mathbf{a}'_{m_1 m_2}|$ for all m_1 and m_2 values. This condition implies that none of the \mathbf{a}' vectors may have the same magnitude as \mathbf{a} . The orthogonality condition (28.17) is automatically fulfilled for these other \mathbf{a}' values, since the integrals will all go to zero. Using now our value $k_1 = 0$, we may attempt to solve Eq. (28.16) with an expansion of the type (28.18).

$$\begin{aligned} \nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 &= \sum \frac{-\omega_{m_1 m_2}^2 + \omega_0^2}{V_0^2} B_{m_1 m_2} e^{-2\pi i(a'_{m_1 m_2} \cdot r)} \\ &= -A \omega_0^2 \sum C_{m_1 m_2} e^{-2\pi i(a'_{m_1 m_2} \cdot r)} \end{aligned} \quad (28.20)$$

where

$$\frac{\omega_{m_1 m_2}^2}{V_0^2} = 4\pi^2 |\mathbf{a}'_{m_1 m_2}|^2 \quad (28.21)$$

We may combine the two sums in Eq. (28.20) and set the coefficients of $e^{-2\pi i(a'_{m_1 m_2} \cdot r)}$ equal to zero. This gives

$$\frac{\omega_0^2 - \omega_{m_1 m_2}^2}{V_0^2} B_{m_1 m_2} = -A \omega_0^2 C_{m_1 m_2} \quad (28.22)$$

and solving for $B_{m_1 m_2}$ in terms of $C_{m_1 m_2}$, we obtain

$$B_{m_1 m_2} = -\frac{A \omega_0^2 V_0^2}{\omega_0^2 - \omega_{m_1 m_2}^2} C_{m_1 m_2} \quad (28.23)$$

We have assumed that $|\mathbf{a}'_{m_1 m_2}| \neq |\mathbf{a}|$, and therefore $\omega_0^2 \neq \omega_{m_1 m_2}^2$, so that $B_{m_1 m_2}$ is always finite. Summarizing the results for case 1,

we see that there is *no first-order correction to the frequency* (since $k_1 = 0$) and the *shape of the waves is slightly perturbed*, in accordance with the general scheme (28.18). Case 1 is characterized by the fact that *one* of the coefficients in A is much larger than the others, which are all of the order of ϵ .

Case 2.—We first discuss the case of an exact coincidence: $\mathbf{a}'' \neq \mathbf{a}$; $\mathbf{a}'' = \mathbf{a}'_{n_1 n_2}$ for some m_1 and m_2 values, say n_1 and n_2 .

In this case then

$$|\mathbf{a}'_{n_1 n_2}| = |\mathbf{a}|, \quad \mathbf{a}'_{n_1 n_2} = \mathbf{a} - n_1 \mathbf{b}_1 - n_2 \mathbf{b}_2 \quad (28.24)$$

This means that there are two values of \mathbf{a} that describe the motion equally well; *i.e.*, there are two solutions for the unperturbed problem that are equivalent. This degenerate motion is the analogue of the motion in the one-dimensional case when $a = \pm 1/2d$. We must write u_0 as a linear combination of two exponentials in this case, one for each of the solutions.

$$u_0 = C e^{-2\pi i(\mathbf{a} \cdot \mathbf{r})} + C' e^{-2\pi i(\mathbf{a}'_{n_1 n_2} \cdot \mathbf{r})} \quad (28.25)$$

with arbitrary coefficients C and C' to be discussed later. The terms in $\mathbf{a}'_{n_1 n_2}$ will turn out to represent the reflected wave in Bragg reflection. Substituting Eq. (28.25) into Eq. (28.9) and using Eq. (28.14) will yield

$$\begin{aligned} \nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 &= -\frac{k_1}{V_0^2} (C e^{-2\pi i(\mathbf{a} \cdot \mathbf{r})} + C' e^{-2\pi i(\mathbf{a}'_{n_1 n_2} \cdot \mathbf{r})}) \\ &\quad - \omega_0^2 \left(\sum_{\substack{m_1 \neq n_1 \\ m_2 \neq n_2}} C C_{m_1 m_2} e^{-2\pi i[(\mathbf{a} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2) \cdot \mathbf{r}]} \right. \\ &\quad \left. + \sum C' C_{m_1 m_2} e^{-2\pi i[(\mathbf{a}'_{n_1 n_2} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2) \cdot \mathbf{r}]} \right) \\ &= - \left[\left(\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1 - n_2} \right) e^{-2\pi i(\mathbf{a} \cdot \mathbf{r})} \right. \\ &\quad \left. + \omega_0^2 \sum_{\substack{m_1 \neq n_1 \\ m_2 \neq n_2}} C' C_{m_1 m_2} e^{-2\pi i[(\mathbf{a}'_{n_1 n_2} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2) \cdot \mathbf{r}]} \right] \\ &\quad - \left[\left(\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1 n_2} \right) e^{-2\pi i(\mathbf{a}'_{n_1 n_2} \cdot \mathbf{r})} \right. \\ &\quad \left. + \omega_0^2 \sum_{\substack{m_1 \neq n_1 \\ m_2 \neq n_2}} C C_{m_1 m_2} e^{-2\pi i(\mathbf{a}'_{m_1 m_2} \cdot \mathbf{r})} \right] \quad (28.26) \end{aligned}$$

The two sums give no trouble since their exponentials are never zero, even when multiplied by $e^{2\pi i(a'',r)}$ in the orthogonality relation (28.17). Therefore, we need consider only terms in $e^{-2\pi i(a,r)}$ and $e^{-2\pi i(a'n_1n_2,r)}$. We use the orthogonality relation to determine k_1 . All terms in the integral (28.17) vanish except the two already noted. We must allow two values for a'' , one corresponding to a and one to $a'_{n_1n_2}$. The first gives the equation

$$\begin{aligned} \int e^{2\pi i(a,r)} \left[\left(\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1-n_2} \right) e^{-2\pi i(a,r)} \right. \\ \left. + \left(\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1n_2} \right) e^{-2\pi i(a'n_1n_2,r)} \right] d\tau \\ = \left(\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1-n_2} \right) \int d\tau = 0 \end{aligned}$$

Since $a \neq a'$ except in magnitude, the second term vanishes on integration. The constant coefficient of the integral must vanish, and this gives one equation in k_1 , C , and C' .

$$\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1-n_2} = 0 \quad (28.27a)$$

The other orthogonality relation is

$$\begin{aligned} \int e^{2\pi i(a'n_1n_2,r)} \left[\left(\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1-n_2} \right) e^{-2\pi i(a,r)} \right. \\ \left. + \left(\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1n_2} \right) e^{-2\pi i(a'n_1n_2,r)} \right] d\tau \\ = \left(\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1n_2} \right) \int d\tau = 0 \end{aligned}$$

by the same reasoning, and this gives a second equation in k_1 , C , and C' .

$$\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1n_2} = 0 \quad (28.27b)$$

These two equations must give the same ratio of C to C' . The condition for this is that the determinant of the coefficients of C and C' vanish. Thus

$$\begin{vmatrix} \frac{k_1}{V_0^2} & \omega_0^2 C_{-n_1-n_2} \\ \omega_0^2 C_{n_1n_2} & \frac{k_1}{V_0^2} \end{vmatrix} = 0 = \begin{vmatrix} \frac{k_1}{V_0^2} & \omega_0^2 C_{n_1n_2}^* \\ \omega_0^2 C_{n_1n_2} & \frac{k_1}{V_0^2} \end{vmatrix} \quad (28.28)$$

since $C_{-n_1-n_2} = C_{n_1n_2}^*$ is necessary to keep f real. Solving Eq. (28.28) for k_1 gives

$$\left(\frac{k_1}{V_0^2}\right)^2 - \omega_0^4 |C_{n_1n_2}|^2 = 0 \quad \text{or} \quad k_1 = \pm \omega_0^2 V_0^2 |C_{n_1n_2}| \quad (28.29)$$

Once k_1 is obtained, the ratio C'/C results from Eq. (28.27), and the remaining coefficients of expansion (28.18) are computed without difficulty.

Summarizing the results for case 2, we obtain a first-order correction on the frequency

$$\omega^2 = \omega_0^2 + \epsilon k_1 = \omega_0^2 \pm \epsilon \omega_0^2 V_0^2 |C_{n_1n_2}| \quad (28.30)$$

and a wave whose general shape is still represented by Eq. (28.18), but which has *two* large coefficients (C and C') in its expansion instead of a single one as in case 1. All other coefficients in the amplitude function $A(\mathbf{r})$ are of the first order in ϵ .

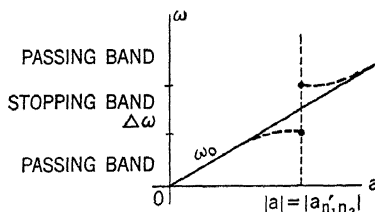


FIG. 28.1.

We have already seen in case 1 that the perturbation is small when \mathbf{a} is not too close to any $\mathbf{a}'_{n_1n_2}$. When it is, however, the perturbation becomes larger and when $\mathbf{a} = \mathbf{a}'_{n_1n_2}$, the frequency has two possible values, one for each of the values of k_1 given by Eq. (28.29). Thus for the unperturbed wave the frequency is a linear function of $|\mathbf{a}|$ as shown by the solid line in Fig. 28.1. The dotted line shows ω as a function of \mathbf{a} for the perturbed wave. For a certain value of \mathbf{a} the curve splits up and the perturbation becomes less as \mathbf{a} gets farther away from $\mathbf{a}'_{n_1n_2}$.

29. The Exceptional Waves of Case 2 and Bragg Reflection

We shall discuss later the transition from the portion of the curve that remains unperturbed and the exceptional point at which the splitting occurs, but first we wish to compare the conditions for case 2 with Bragg's formula for reflection from crystal planes. Bragg discovered that X rays can be selectively

reflected from some of the atomic planes in crystal lattices, provided a certain relation among wave length, the distance of separation of the planes, and the direction of propagation is satisfied. We want to prove that this condition is equivalent to

$$|\mathbf{a}| = |\mathbf{a}_{n_1 n_2}| \quad (29.1)$$

for the (n_1, n_2) values that define the atomic planes (which in the two-dimensional case are just rows of atoms). This is, of course, exactly the condition for case 2 of Sec. 28.

We consider the vector in the reciprocal lattice

$$\mathbf{B} = n_1 \mathbf{b}_1 + n_2 \mathbf{b}_2$$

This vector defines a direction in the two-dimensional reciprocal lattice and a row of atoms perpendicular to this direction in the plane of the direct lattice. This row is the row responsible for the Bragg reflection. For the reflection to take place, Eq. (29.1) must hold. We denote the \mathbf{a} for which Bragg's condition holds by \mathbf{a}_0 and replace $\mathbf{a}'_{n_1 n_2}$ by \mathbf{a}'_0 . Thus

$$\mathbf{a}'_0 = \mathbf{a}_0 - \mathbf{B} \quad (29.2)$$

and

$$|\mathbf{a}'_0| = |\mathbf{a}_0| \quad (29.1)$$

In other words,

$$|\mathbf{a}'_0|^2 = |\mathbf{a}_0|^2 = |\mathbf{a}_0 - \mathbf{B}|^2 = |\mathbf{a}_0|^2 - 2(\mathbf{a}_0 \cdot \mathbf{B}) + |\mathbf{B}|^2$$

and

$$|\mathbf{a}_0|^2 = |\mathbf{a}'_0|^2 = |\mathbf{a}'_0 + \mathbf{B}|^2 = |\mathbf{a}'_0|^2 + 2(\mathbf{a}'_0 \cdot \mathbf{B}) + |\mathbf{B}|^2$$

These two equations yield

$$|\mathbf{B}|^2 = 2(\mathbf{a}_0 \cdot \mathbf{B}) = -2(\mathbf{a}'_0 \cdot \mathbf{B})$$

or

$$|\mathbf{a}_0| \cos(\mathbf{a}_0, \mathbf{B}) = -|\mathbf{a}'_0| \cos(\mathbf{a}'_0, \mathbf{B}) = \frac{1}{2}|\mathbf{B}| \quad (29.3)$$

and since \mathbf{a}_0 and \mathbf{a}'_0 differ only in direction but have the same length,

$$\cos(\mathbf{a}_0, \mathbf{B}) = -\cos(\mathbf{a}'_0, \mathbf{B})$$

or \mathbf{a}_0 makes the same angle with \mathbf{B} that \mathbf{a}'_0 makes with $-\mathbf{B}$. From Eq. (29.3) it now follows at once that \mathbf{a}_0 terminates on the perpendicular bisector of \mathbf{B} , while \mathbf{a}'_0 terminates on the perpendicular bisector of $-\mathbf{B}$. This means that the projection of \mathbf{a} on

\mathbf{B} has magnitude $|\mathbf{B}|/2$ or

$$|a| \cos \varphi = \frac{1}{\lambda} \cos \varphi = \frac{|\mathbf{B}|}{2} = \frac{1}{2\delta}$$

or

$$\lambda = 2\delta \cos \varphi \quad (29.4)$$

where φ is the angle between \mathbf{a} and \mathbf{B} and δ is the distance between successive rows passing through the lattice points in the direct lattice and perpendicular to the basis vector conjugate to \mathbf{B} [Eq. (25.22)]. This is illustrated in Fig. 29.1, showing a set of parallel rows in the direct lattice, with a distance of separation δ , with the incident beam a_0 , and with the reflected beam a'_0 . The elementary theory of Bragg reflection is as follows:

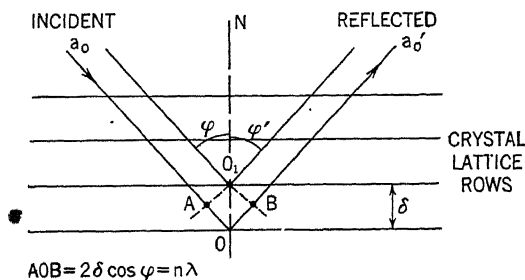


FIG. 29.1.—Bragg's reflection.

1. Angles φ and φ' must be equal, ensuring a uniform reflection from each lattice row.

2. Waves reflected from two successive lattice rows must be in phase; hence

$$AOB = 2\delta \cos \varphi = m\lambda \quad (29.5)$$

This is Bragg's formula, which checks with Eq. (29.4) when $m = 1$. Let us specify the integers defining the particular vector \mathbf{B} corresponding to $m = 1$ by n_1 and n_2 . We note that other vectors \mathbf{B}_m of the reciprocal lattice are obtained if we take

$$\mathbf{B}_m = m\mathbf{B} = mn_1\mathbf{b}_1 + mn_2\mathbf{b}_2$$

and yield rows in the direct lattice separated by a distance

$$\delta_m = \frac{1}{|\mathbf{B}_m|} = \frac{1}{m|\mathbf{B}|} = \frac{\delta}{m}$$

from which we obtain the generalization contained in Eq. (29.5).

All this proves that case 2, with its two large components in the wave, corresponds exactly to Bragg's condition for reflection; *i.e.*, to the situation in which an incident C wave (\mathbf{a}_0) can be reflected from atomic rows in the lattice and generate a C' wave (\mathbf{a}'_0) of large amplitude.

30. Transition near the Discontinuity

We wish now to investigate the region where \mathbf{a} almost satisfies Bragg's condition but not quite; *i.e.*, we allow

$$\mathbf{a} = \mathbf{a}_0 + \eta\mathbf{B} \quad \text{where } \eta \text{ is small} \quad (30.1)$$

Then

$$\mathbf{a}' = \mathbf{a} - \mathbf{B} = \mathbf{a}_0 - \mathbf{B} + \eta\mathbf{B} = \mathbf{a}'_0 + \eta\mathbf{B} \quad (30.2)$$

where \mathbf{a}_0 and \mathbf{a}'_0 are the vectors defined in Secs. 28 and 29 and exactly satisfy Bragg's reflection condition. The new vectors

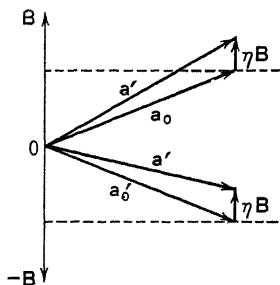


FIG. 30.1.

\mathbf{a} and \mathbf{a}' satisfy these same conditions approximately only, according to the η terms. This is shown in Fig. 30.1. The squares of the absolute values of \mathbf{a} and \mathbf{a}' are given by

$$\left. \begin{aligned} |\mathbf{a}|^2 &= |\mathbf{a}_0|^2 + 2\eta(\mathbf{a}_0 \cdot \mathbf{B}) \\ |\mathbf{a}'|^2 &= |\mathbf{a}'_0|^2 + 2\eta(\mathbf{a}'_0 \cdot \mathbf{B}) = |\mathbf{a}_0|^2 - 2\eta(\mathbf{a}_0 \cdot \mathbf{B}) \end{aligned} \right\} \quad (30.3)$$

according to Eq. (29.3) and by dropping η^2 term. Furthermore, we define the following expressions:

$$\left. \begin{aligned} \frac{\omega_0^2}{V_0^2} &= 4\pi^2|\mathbf{a}_0|^2 \\ \frac{\omega_a^2}{V_0^2} &= 4\pi^2|\mathbf{a}|^2 = \frac{\omega_0^2}{V_0^2} + 8\pi^2\eta(\mathbf{a}_0 \cdot \mathbf{B}) \\ \frac{\omega_{a'}^2}{V_0^2} &= 4\pi^2|\mathbf{a}'|^2 = \frac{\omega_0^2}{V_0^2} - 8\pi^2\eta(\mathbf{a}_0 \cdot \mathbf{B}) \end{aligned} \right\} \quad (30.4)$$

Also

$$u_0 = Ce^{-2\pi i(a \cdot r)} + C'e^{-2\pi i(a' \cdot r)} \quad (30.5)$$

as in Eq. (28.25), and

$$\begin{aligned} u &= u_0 + \epsilon u_1 \\ \frac{1}{V^2} &= \frac{1}{V_0^2} + \epsilon f \end{aligned} \quad (30.6)$$

The frequency ω of the u_0 wave is given by

$$\omega^2 = \omega_0^2 + \epsilon k_1$$

as in Eq. (28.7). This frequency ω is the same for both terms entering u_0 and should not be confused with ω_a and $\omega_{a'}$, which represent the frequencies of the (C, C') waves in the unperturbed medium, while ω is the common frequency of both terms in the medium with periodic perturbation.

We start again from Eq. (28.8), and we take Eq. (30.4) into account. The $\nabla^2 u_0$ term yields

$$\begin{aligned} \nabla^2 u_0 &= \nabla^2 (Ce^{-2\pi i(a \cdot r)} + C'e^{-2\pi i(a' \cdot r)}) \\ &= -4\pi^2 |\mathbf{a}|^2 Ce^{-2\pi i(a \cdot r)} - 4\pi^2 |\mathbf{a}'|^2 C'e^{-2\pi i(a' \cdot r)} \\ &= -\frac{\omega_0^2}{V_0^2} u_0 - 8\pi^2 \eta (\mathbf{a}_0 \cdot \mathbf{B}) (Ce^{-2\pi i(a \cdot r)} - C'e^{-2\pi i(a' \cdot r)}) \end{aligned} \quad (30.7)$$

The new η terms characterize our present problem. η and ϵ are both perturbation coefficients and are assumed to be of the same order of magnitude. We set

$$h = \frac{\eta}{\epsilon} \quad (30.8)$$

We now separate the perturbation terms from the rest of the equation and set them equal to zero as before.

$$\begin{aligned} \nabla^2 u_1 + \frac{\omega_0^2}{V_0^2} u_1 &= - \left(\frac{k_1}{V_0^2} + \omega_0^2 f \right) (Ce^{-2\pi i(a \cdot r)} + C'e^{-2\pi i(a' \cdot r)}) \\ &\quad + 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) (Ce^{-2\pi i(a \cdot r)} - C'e^{-2\pi i(a' \cdot r)}) \end{aligned} \quad (30.9)$$

This is exactly Eq. (28.9) but for the additional h term. The discussion from now on parallels that given in Sec. 28, case 2. We again expand f according to Eq. (28.14) and group the terms that contribute to the perturbation as in Eq. (28.26). The two important terms are

$$\text{and } \left. \begin{aligned} & \left[\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1 - n_2} - 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) C \right] e^{-2\pi i (\mathbf{a} \cdot \mathbf{r})} \\ & \left[\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1 n_2} + 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) C' \right] e^{-2\pi i (\mathbf{a}' \cdot \mathbf{r})} \end{aligned} \right\} \quad (30.10)$$

When we substitute the right-hand side of Eq. (30.9) in the orthogonality relation (28.17), all terms contribute zero to the integral except the two above. The first will give rise to one equation similar to Eq. (28.27a) for C and C' when the solution $e^{-2\pi i (\mathbf{a} \cdot \mathbf{r})}$ of the homogeneous equation for u_1 is used.

$$\frac{k_1 C}{V_0^2} + \omega_0^2 C' C_{-n_1 - n_2} - 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) C = 0 \quad (30.11a)$$

The second one gives another relation similar to Eq. (28.27b), when $u_1 = e^{-2\pi i (\mathbf{a}' \cdot \mathbf{r})}$ is used.

$$\frac{k_1 C'}{V_0^2} + \omega_0^2 C C_{n_1 n_2} + 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) C' = 0 \quad (30.11b)$$

The determinant of the coefficients of C and C' must vanish as before. This determinant is

$$\begin{vmatrix} \frac{k_1}{V_0^2} - 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) & \omega_0^2 C_{n_1 n_2}^* \\ \omega_0^2 C_{n_1 n_2} & \frac{k_1}{V_0^2} + 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) \end{vmatrix} = 0 \quad (30.12)$$

Solving for k_1 yields

$$\frac{k_1}{V_0^2} = \pm \sqrt{\omega_0^4 |C_{n_1 n_2}|^2 + [8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B})]^2} \quad (30.13)$$

For $h = 0$ we have the Bragg condition and the same results as before. When h is large,

$$\frac{k_1}{V_0^2} = \pm 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) \quad h \gg 1 \quad (30.14)$$

But

$$\frac{\omega^2}{V_0^2} = \frac{\omega_0^2}{V_0^2} + \epsilon \frac{k_1}{V_0^2} = \frac{\omega_0^2}{V_0^2} \pm \epsilon 8\pi^2 h (\mathbf{a}_0 \cdot \mathbf{B}) = \begin{cases} \frac{\omega_a^2}{V_0^2} \\ \frac{\omega_{a'}^2}{V_0^2} \end{cases}$$

according to Eq. (30.4). Thus \mathbf{a} will be in one zone and \mathbf{a}' in

another, and the values for ω will lie on the dotted transition curve shown in Fig. 28.1.

The whole discussion of Secs. 28 and 30 shows that discontinuities in the function $\nu(\mathbf{a})$ appear only when the \mathbf{a} vector has its extremity on the perpendicular bisector of the vectors in the reciprocal lattice. This justifies the rule given in Sec. 27 for the construction of successive zones. Some examples of the shape of the first zone have been found by various authors, and the general rule was given by the present writer, with a comprehensive discussion of the structure of higher zones in two and three dimensions.

31. Examples and Discussion of Zones in Two Dimensions

The theory just developed in Secs. 28, 29, and 30 contains the principle of X-ray reflection from crystal lattices. X rays propagate through the crystal with the velocity of light in vacuum, and atoms or molecules may just slightly perturb the propagation. The perturbation is practically proportional to the electronic density, and an equation of propagation of the type of Eq. (28.8) is obtained, with a perturbation term f proportional to the electronic density. In a crystal lattice, atoms are regularly distributed along a direct lattice, each atomic nucleus being surrounded by a cloud of electrons that may partly overlap that of its neighbors. The electronic density is a periodic function with the periodicity of the lattice and can be expanded in a multiple Fourier series like Eq. (28.14). In the final results given by Eq. (28.30) or Eq. (30.13) only the absolute value $|C_{m_1 m_2}|$ of the coefficients of the Fourier terms appeared. Hence, any experiment on wave propagation through the crystal lattice will give only the absolute value and not the phase angle. This latter may often be obtained from symmetry considerations and some general knowledge of the lattice structure, and then the whole Fourier expansion of the electronic space charge is found, from which the distribution of the space charge at any point can be computed. Figure 31.1 gives an example of such experimental results.

This is the principle of the procedure announced at the end of Sec. 26. As a rule, *all* coefficients C_m of the Fourier expansion are required, and there is practically no example of actual lattices where simplified assumptions such as Eq. (26.1) or (26.8) could be used. Sometimes a few exceptional terms may happen to be zero, but this is not very frequent.

The relation between the conditions yielding discontinuities in the $\nu(a)$ curve and Bragg's reflection condition was explained in Sec. 29. In principle the explanation is the same as for the relation between passing bands and stopping bands in the one-dimensional lattices discussed in Sec. 16: Let a certain direction

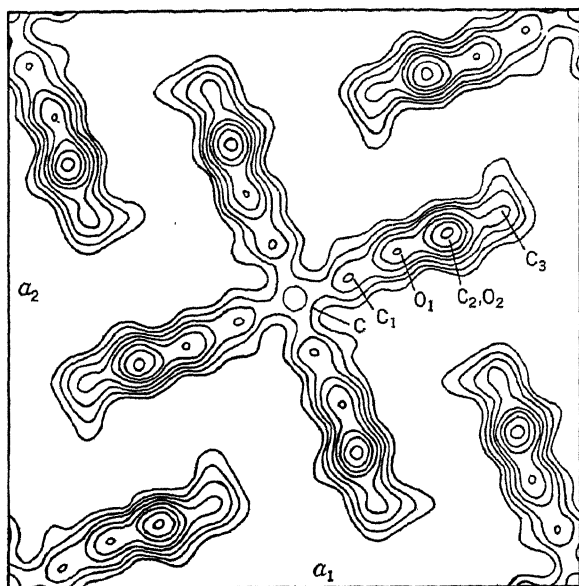


FIG. 31.1.—Crystal structure of pentaerythritol tetracetate $C \cdot (CH_2O \cdot CO \cdot CH_3)_4$ showing the position of carbon and oxygen atoms as resulting from a Fourier analysis based on X-rays diffraction. (Courtesy of T. H. Goodwin and R. Hardy, *Proc. Roy. Soc. (London) A*, vol. 164 (1938), p. 369.)

of propagation be given (direction of a) and the frequency ν of the wave be varied. If the frequency corresponds to a passing band, then the wave falling upon the lattice with this direction of propagation and frequency may be propagated through the crystal. Surface conditions at the boundary of the crystal will give only partial reflection from the surface. If, however, the frequency ν falls inside one of the stopping bands, like the intervals obtained in Fig. 28.1, then the corresponding wave cannot be propagated through the crystal and must be totally reflected from the surface.

The elementary Bragg theory, as sketched in Fig. 29.1, predicted reflection for just one frequency, while our more comprehensive treatment yields reflection for the whole stopping

band $\Delta\nu$. For X rays, the perturbation of the wave by the crystal lattice is extremely small and the $\Delta\nu$ bands for ideal crystals are very narrow.

All this was explained here for a two-dimensional lattice structure, but X-ray reflection actually occurs in three-dimensional structures. Electron reflection according to wave mechanics (electrons being represented by De Broglie waves) is a more

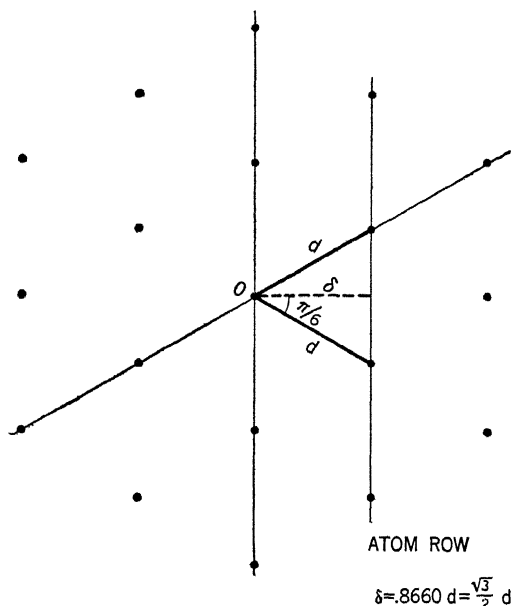


FIG. 31.2.—Hexagonal lattice.

accurate example of a two-dimensional problem, since in many experiments (Davisson and Germer) the reflection takes place on the surface of the crystal (two dimensions) and electronic waves do not penetrate inside the crystal.

We now give a few examples of direct and reciprocal lattices in two dimensions, with the corresponding zone structures. Where a certain direct lattice is given, the shortest way to find the reciprocal lattice is to use Eq. (25.22) and look for some parallel rows of lattice points in the direct lattice, compute their distance of separation δ , and obtain a vector of the reciprocal lattice by taking a length $1/\delta$ in the direction perpendicular to the rows. This is shown in Fig. 31.2 for a hexagonal lattice based on two

vectors \mathbf{d} making an angle $\pi/3$. The reciprocal lattice is another hexagonal lattice, turned through an angle $\pi/6$ and with vectors \mathbf{b} .

$$|\mathbf{b}| = \frac{2}{\sqrt{3}} \frac{1}{|\mathbf{d}|}$$

as shown in Fig. 31.3. The first five zones have been drawn and

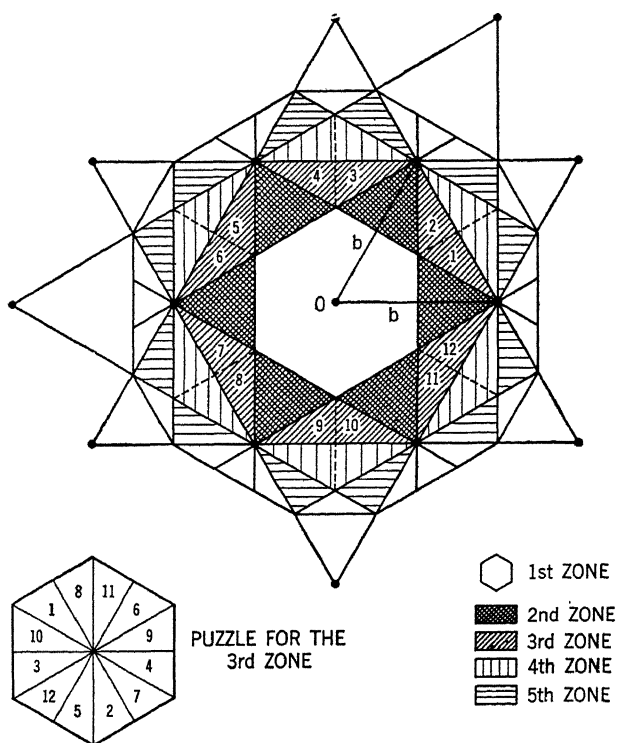


FIG. 31.3.—Zones for a hexagonal lattice.

the sixth one is easily recognized. Each zone covers an area equal to that of the first one or to that of the (\mathbf{b}, \mathbf{b}) parallelogram.

$$A_b = |\mathbf{b}|^2 \sin \frac{\pi}{3} = \frac{\sqrt{3}}{2} |\mathbf{b}|^2 = \frac{2}{\sqrt{3}} \frac{1}{|\mathbf{d}|^2}$$

while the area of the cell in the direct lattice is

$$A_d = |\mathbf{d}|^2 \sin \frac{\pi}{3} = \frac{\sqrt{3}}{2} |\mathbf{d}|^2 = \frac{1}{A_b}$$

Each zone can be reduced to the first zone by taking its sections and giving them a translation parallel and equal to one of the vectors of the reciprocal lattice. This is obvious for the second zone in Fig. 31.3. The case of the third zone is illustrated by a mosaic showing how the different sections of the third zone exactly cover the first one after being given the necessary translation. Similar mosaics can be drawn for higher zones.

Oblique lattices were used in Figs. 25.1, 27.2, and 27.4. The case of a rectangular lattice is shown in Fig. 31.4, where the first four zones have been drawn, each of which has an area equal to

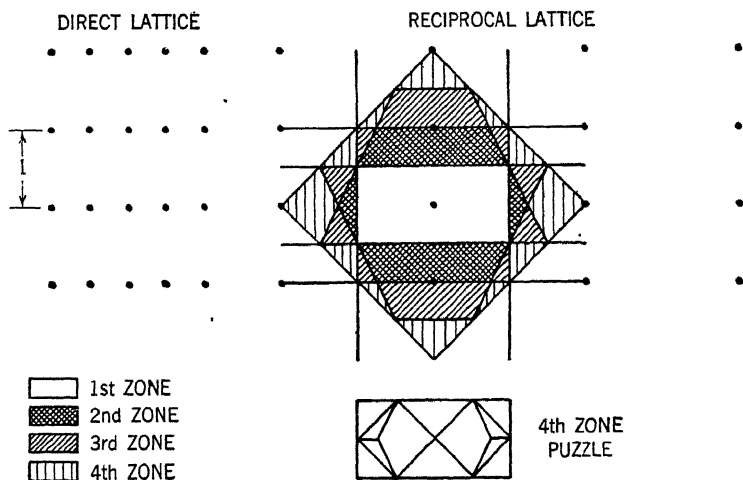


FIG. 31.4.—Rectangular lattice.

that of the first one. Translation of the different sections of one zone by vectors of the reciprocal lattice can be used to superpose them on the first zone. The corresponding mosaic is shown for the fourth zone. Figures 31.5 and 31.6 refer to the square-lattice structure and are original drawings given by the writer in a paper published in 1930. Figure 31.6 contains the mosaics for the successive zones in the square lattice.

All this theory is based upon the assumption of the periodicity of the function f of Sec. 28. It is not inconceivable that in certain physical problems the perturbation function f (the electronic density, for instance) may have a lower symmetry than the atomic lattice, and that its periodicity may offer a different character. In such cases the zone structure would

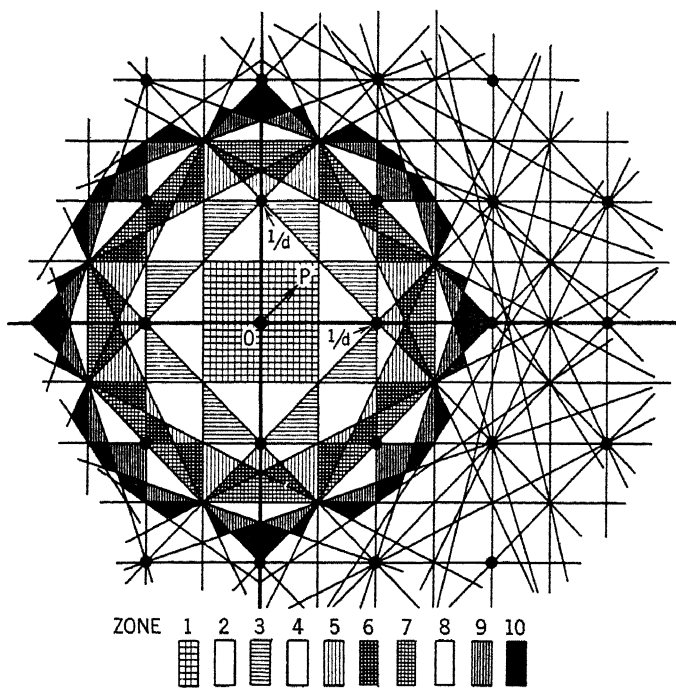


FIG. 31.5.—Zones for a square lattice.

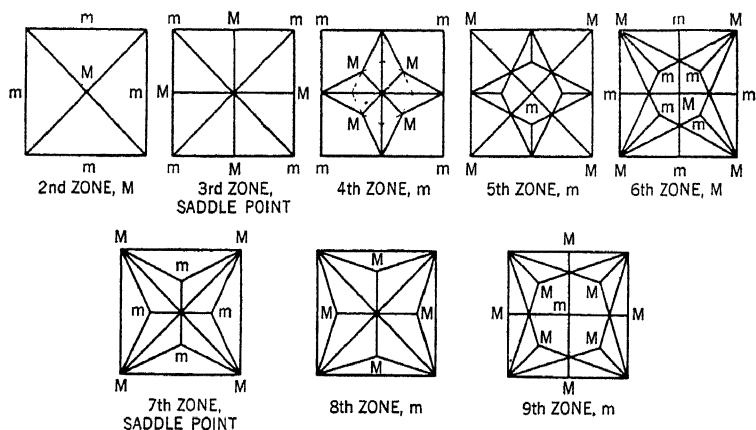


FIG. 31.6.—Square lattice, reduction of the first zones.

correspond to the actual periodicity and symmetry of the perturbation function and not to those of the lattice itself. This may be the case when the individual electronic clouds around each atomic nucleus already possess a certain structure (atoms in P , D , . . . states) and do not exhibit spherical symmetry (S state) for the isolated atoms.

Once the zonal structure has been obtained, most of the results established in the first chapters for one-dimensional structures can be readily extended to two-dimensional problems. Zones

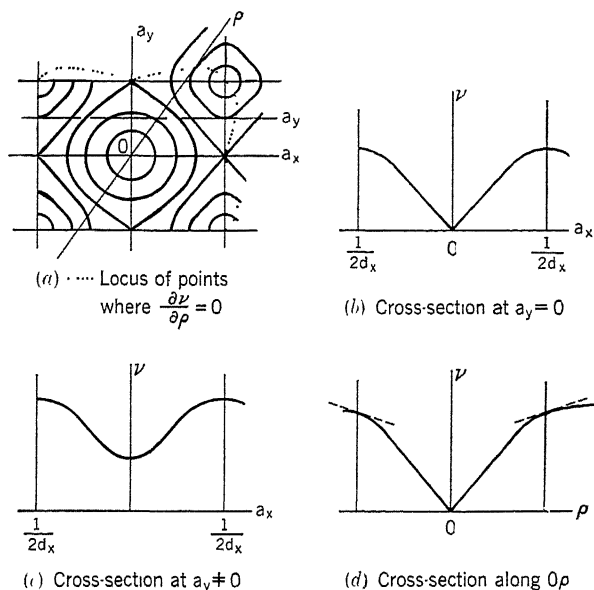


FIG. 31.7.

1, 2, 3, . . . correspond to intervals of similar numbers in one dimension, as explained in Fig. 27.1. The frequency ν was a periodic function of period $1/d$ in one dimension, and continuity across the boundary of the first zone meant that the $\nu(a)$ curve should reach this boundary with a horizontal tangent. In two dimensions $\nu(a)$ is a function of the two variables a_x and a_y (components of \mathbf{a}), and it must reach the boundary of the first zone with a zero normal derivative. We can plot a map of ν inside the zone, using lines of equal ν values (like lines of equal altitude on a map), and obtain a drawing like the one represented in Fig. 31.7a. A cross section of the map along Ox yields a

curve similar to those of Chaps. I, II, and III (Fig. 31.7*b*). Another cross section at a_{y_0} gives the curve of Fig. 31.7*c* with horizontal tangents in the middle and at the ends of the interval. A radial cross section along $O\rho$, as shown in Fig. 31.7*d*, shows no horizontal tangents at the ends of the interval. The distinction between the radial derivative $\partial\nu/\partial\rho$ or a normal derivative is also exemplified in Fig. 31.7*a*, where the normal derivative is zero along the border of the zone, while the locus of point $\partial\nu/\partial\rho = 0$ is represented by the dotted curve. One must always be very careful not to confuse these two different definitions.

Figure 31.7 would represent the qualitative behavior of a *monatomic lattice* for elastic vibrations in the case of a rectangular lattice (d_x, d_y), and it corresponds more closely to problems of

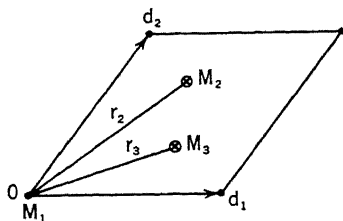


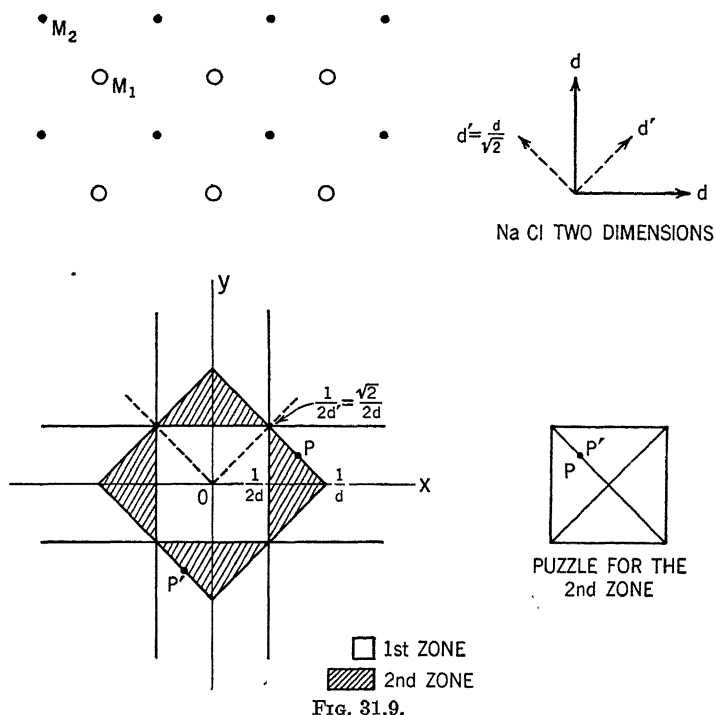
FIG. 31.8.—Lattice with basis.

the type discussed in Chap. III with a low passing band. The phase velocity in the x and y directions is different, and the limiting frequency (cutoff) varies all along the boundary of the first zone. It depends upon the direction of propagation as well as the lattice structure.

A *polyatomic lattice* is one containing several atoms in the fundamental cell, as shown schematically in Fig. 31.8. It is often called a *lattice with basis*, where the word “basis” is used for the bundle of vectors $r_2, r_3, r_4, \dots, r_n$ defining the positions of particles M_2, M_3, \dots, M_n of the cell with respect to a certain particle M_1 taken as the origin. This is the two-dimensional generalization of Fig. 17.1. Such problems resulted, in the one-dimensional case, in a $\nu(a)$ curve with N branches, one acoustical and the remaining $(N - 1)$ optical. Here we obtain N surfaces covering the first zone. The cross section of these N surfaces along a given direction is very similar to the curves in Figs. 3.2 and 3.9 in one dimension.

The NaCl problem may be discussed more accurately in con-

section with the theory developed in Chap. IV. When the particles M_1 and M_2 differ in mass, the lattice is a square-centered lattice as shown in Fig. 31.9 and represents a lattice with basis, built upon two equal orthogonal vectors \mathbf{d} and \mathbf{d} . Its first two zones are shown at the bottom of the drawing. When, however, $M_1 = M_2$, the lattice suddenly changes its character and becomes a simple square lattice, built on two vectors $\mathbf{d}' = \mathbf{d}/\sqrt{2}$ at 45 deg. Its first zone includes both the



first and the second zones of the previous structure. In this case ($M_1 = M_2$) we obtain $\nu(\mathbf{a})$ as a single-valued function of \mathbf{a} over the whole big zone. When $M_1 \neq M_2$, the four outer triangular sections (second zone) must be translated to the reduced first zone, and $\nu(\mathbf{a})$ becomes a double-valued function of \mathbf{a} inside this new first zone. The limits of these zones are at $\pm 1/2d$ and $\pm 1/d$ along the x or the y axis (as in the one-dimensional problem) except for directions at 45 deg.; here the limits of the zones coincide. The mosaic of the sections of the second

zone shows how two points P and P' come in contact with the new representation. Since ν obviously has the same values at P and P' , by symmetry there is *no discontinuity* of $\nu(\mathbf{a})$ along the diagonals of the mosaic. Both branches of the $\nu(\mathbf{a})$ function are continuous except at the boundary of the new first zone.

N particles per cell mean N branches in the $\nu(\mathbf{a})$ function when \mathbf{a} is restricted to the interior of the first zone. It may also be thought of as an extension of one branch $\nu(\mathbf{a})$ over N zones, as in the preceding paragraph.

A continuous periodic structure is obtained at the limit $N \rightarrow \infty$ and yields an infinite number of branches inside the first zone or a one-branch function extending over the whole plane. This is the problem studied in Secs. 28 and 30. For a uniform medium with constant properties, the one-branch function

$$\nu = W|\mathbf{a}|$$

extends all over the \mathbf{a} plane. Periodic perturbation introduces discontinuities in the function $\nu(\mathbf{a})$ on all lines chosen as the limits of a zone (Secs. 27 and 29) and corresponding to Bragg reflection. A large perturbation simply increases the discontinuities without ever changing their location in the \mathbf{a} plane. All the *sections* of an arbitrary zone can be brought back to the first zone by translations $h_1\mathbf{b}_1 + h_2\mathbf{b}_2$ in the *reciprocal lattice* and fit into a mosaic just covering the first zone. When this is done, each zone yields one branch of the $\nu(\mathbf{a})$ function reduced inside the first zone.

As stated before, experimental evidence shows that an actual crystal lattice exhibits the *complete* row or plane *system*, with no set missing. Therefore, the complete Fourier expansion of Sec. 26 and the complete system of discontinuity lines or zone limits must occur.

CHAPTER VII

THREE-DIMENSIONAL LATTICES

32. Direct and Reciprocal Lattices in Three Dimensions

The lattice in three dimensions is usually defined by three oblique coordinates. The elementary cell is thus a parallelepiped. We take the basis vectors \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_3 to be the vectors joining the origin of the lattice with three particles in the lattice that do not all lie in the same plane. These vectors define the first elementary cell. Then any other set of basis vectors¹ is defined by

$$\left. \begin{aligned} \mathbf{d}'_1 &= \alpha_{11}\mathbf{d}_1 + \alpha_{12}\mathbf{d}_2 + \alpha_{13}\mathbf{d}_3 \\ \mathbf{d}'_2 &= \alpha_{21}\mathbf{d}_1 + \alpha_{22}\mathbf{d}_2 + \alpha_{23}\mathbf{d}_3 \\ \mathbf{d}'_3 &= \alpha_{31}\mathbf{d}_1 + \alpha_{32}\mathbf{d}_2 + \alpha_{33}\mathbf{d}_3 \end{aligned} \right\} \begin{array}{l} \alpha_{ik} = \text{positive or} \\ \text{negative integers} \end{array} \quad (32.1)$$

where the determinant of the α_{ik} must not be zero to ensure that \mathbf{d}'_1 , \mathbf{d}'_2 , and \mathbf{d}'_3 are not linearly dependent; *i.e.*, \mathbf{d}'_1 , \mathbf{d}'_2 , and \mathbf{d}'_3 do not all lie in the same plane. A lattice point is given by the vector

$$\mathbf{R}_{l_1 l_2 l_3} = l_1 \mathbf{d}_1 + l_2 \mathbf{d}_2 + l_3 \mathbf{d}_3 \quad (32.2)$$

where l_1 , l_2 , and l_3 are integers. Points in the first elementary cell have vectors of the form

$$\mathbf{r} = \xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2 + \xi_3 \mathbf{d}_3 \quad (32.3)$$

where $|\xi_i| < 1$. Any other point in the lattice may be obtained by adding vectorially the vector for the origin of the cell in which the point lies to its vector from the origin of its cell.

The *reciprocal lattice* is defined in exactly the same fashion as for two dimensions: Its basis vectors \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 satisfy the nine equations analogous to Eq. (25.3).

$$(\mathbf{b}_i \cdot \mathbf{d}_k) = \delta_{ik} \quad (32.4)$$

The propagation vector for a wave propagating in the direct lattice is drawn in the reciprocal lattice as before. The analogue

¹ See footnote on p. 95.

to the fact that direct and reciprocal cells have reciprocal areas in two dimensions [Eq. (25.14)] is that they have reciprocal volumes in three dimensions.

$$V_d V_b = 1 \quad (32.5)$$

To prove this, we note that \mathbf{b}_3 is perpendicular to \mathbf{d}_1 and \mathbf{d}_2 from Eq. (32.4). Thus we may write

$$\mathbf{b}_3 = k(\mathbf{d}_1 \times \mathbf{d}_2) \quad (32.6)$$

Now

$$(\mathbf{b}_3 \cdot \mathbf{d}_3) = 1 = k[\mathbf{d}_3 \cdot (\mathbf{d}_1 \times \mathbf{d}_2)] = kV_d \quad (32.7)$$

Therefore, $k = 1/V_d$, and from Eq. (32.6) it follows that

$$\mathbf{b}_3 = \frac{1}{V_d} (\mathbf{d}_1 \times \mathbf{d}_2) \quad (32.8)$$

Similarly,

$$\left. \begin{aligned} \mathbf{b}_1 &= \frac{1}{V_d} (\mathbf{d}_2 \times \mathbf{d}_3) \\ \mathbf{b}_2 &= \frac{1}{V_d} (\mathbf{d}_3 \times \mathbf{d}_1) \end{aligned} \right\} \quad (32.9)$$

and conversely

$$\begin{aligned} \mathbf{d}_3 &= \frac{1}{V_b} (\mathbf{b}_1 \times \mathbf{b}_2), & \mathbf{d}_1 &= \frac{1}{V_b} (\mathbf{b}_2 \times \mathbf{b}_3), \\ & & \mathbf{d}_2 &= \frac{1}{V_b} (\mathbf{b}_3 \times \mathbf{b}_1) \end{aligned} \quad (32.10)$$

Therefore,

$$(\mathbf{b}_3 \cdot \mathbf{d}_3) = 1 = \frac{[(\mathbf{d}_1 \times \mathbf{d}_2) \cdot (\mathbf{b}_1 \times \mathbf{b}_2)]}{V_d V_b}$$

and if we break the vectors in the numerator up into their Cartesian components and rearrange them, it becomes

$$(\mathbf{d}_1 \cdot \mathbf{b}_1)(\mathbf{d}_2 \cdot \mathbf{b}_2) - (\mathbf{d}_2 \cdot \mathbf{b}_1)(\mathbf{d}_1 \cdot \mathbf{b}_2) = 1$$

so that we obtain the desired relation. *

Another theorem that we shall find useful is that the position vector of any lattice point in the reciprocal lattice is perpendicular to an infinite set of lattice planes in the direct lattice [the two-dimensional analogue is expressed by Eq. (25.22)] where a lattice plane is defined as a plane passing through a set of lattice points. Thus, if we denote the position vector of a

lattice point in the reciprocal lattice by

$$\mathbf{H} = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3 \quad (32.11)$$

where h_1 , h_2 , and h_3 are integers, and let

$$\mathbf{n} = \frac{\mathbf{H}}{|\mathbf{H}|} \quad (32.12)$$

be the unit vector in the direction \mathbf{H} , the vector equation representing a plane perpendicular to the vector \mathbf{H} is

$$(\mathbf{n} \cdot \mathbf{r}) = \frac{(\mathbf{H} \cdot \mathbf{r})}{|\mathbf{H}|} = C \quad (32.13)$$

where C represents the distance of the plane from the origin. Such a plane passes through a point $\mathbf{R}_{l_1 l_2 l_3}$ of the direct lattice if

$$\frac{\mathbf{H} \cdot \mathbf{R}_{l_1 l_2 l_3}}{|\mathbf{H}|} = \frac{h_1 l_1 + h_2 l_2 + h_3 l_3}{|\mathbf{H}|} = C \quad (32.14)$$

by direct substitution for \mathbf{H} and $\mathbf{R}_{l_1 l_2 l_3}$ from Eqs. (32.2) and (32.11). The numerator is an integer m . Now the distance between these lattice planes in the direct lattice is given by the smallest allowable variation of the right-hand member of Eq. (32.14). Since the numerator must be integral,¹ its smallest value is 1 and the distance:

$$\delta = \frac{1}{|\mathbf{H}|} \quad (32.15)$$

if Eq. (32.14) is reduced to lowest terms. One may also state this theorem by saying that the distance between the lattice planes perpendicular to a given direction \mathbf{n} is the reciprocal of the distance of the lattice point in the reciprocal lattice nearest to the origin and in the direction \mathbf{n} from the origin. An analogous theorem may be stated and proved in the reciprocal lattice.

The density of points in the lattice planes is proportional to $\delta = \frac{1}{|\mathbf{H}|}$. For the volume density is constant, and hence the number of points per unit area of the plane is directly proportional to δ . This result is, of course, also true for reciprocal lattices with appropriate changes in notation.

A vector in the direct lattice may be written

$$\mathbf{r} = \xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2 + \xi_3 \mathbf{d}_3 \quad (32.16)$$

¹ See last sentence on p. 98.

where we have the relations

$$\xi_1 = (\mathbf{b}_1 \cdot \mathbf{r}), \quad \xi_2 = (\mathbf{b}_2 \cdot \mathbf{r}), \quad \xi_3 = (\mathbf{b}_3 \cdot \mathbf{r})$$

as may easily be verified. Similarly, for a vector in the reciprocal lattice

$$\mathbf{H} = \eta_1 \mathbf{b}_1 + \eta_2 \mathbf{b}_2 + \eta_3 \mathbf{b}_3 \quad (32.17)$$

we have

$$\eta_1 = (\mathbf{d}_1 \cdot \mathbf{H}), \quad \eta_2 = (\mathbf{d}_2 \cdot \mathbf{H}), \quad \eta_3 = (\mathbf{d}_3 \cdot \mathbf{H})$$

Before going on to the propagation of waves, it is well to give some examples of three-dimensional lattices and their reciprocals.

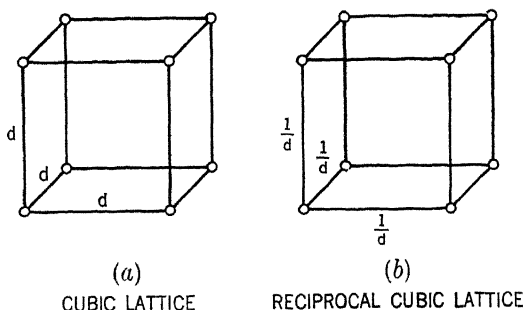


FIG. 32.1.

The simplest three-dimensional lattice structure is probably the *cubic lattice* shown in Fig. 32.1a. Here

$$(\mathbf{d}_i \cdot \mathbf{d}_k) = \begin{cases} d^2 & i = k \\ 0 & i \neq k \end{cases}$$

where d is the length of one edge of the cube. From Eq. (32.9) we obtain the reciprocal lattice $|\mathbf{b}_i| = 1/d$ and \mathbf{b}_i parallel to \mathbf{d}_i . This reciprocal lattice, shown in Fig. 32.1b, is also cubic.

The simple cubic lattice does not occur naturally. There are, however, some related lattices that do. Let us consider the *face-centered cubic lattice* shown in Fig. 32.2a. This lattice has atoms in the center of each of the faces of the cube as well as at the corners. We think of the *elementary cell* in this lattice as the parallelepiped determined by the lines AE , AF , and AG . The volume of this cell is $V_d = d^3/4$ if the edge of the cube is d . The cell based on these vectors contains only one atom A , all others being considered as belonging to similar adjacent cells. It is easily verified that linear combinations of these three fundamental vectors give all the points in the lattice. The sum

of the three vectors leads to B , while $AE + AF - AG$ yields point D . The cube d cannot be used as a fundamental cell, since it contains four atoms instead of one. If the atoms located in the faces of the cube were different from those at the corners, the lattice would become a cubic lattice with a basis $A EFG$, but this point of view is not logical for equal particles.

To obtain the reciprocal of the face-centered cubic lattice we make use of the theorem that a vector in the reciprocal lattice drawn to the point nearest to the origin in a given direction is perpendicular to a set of lattice planes in the direct lattice and has a magnitude equal to the reciprocal of the distance between

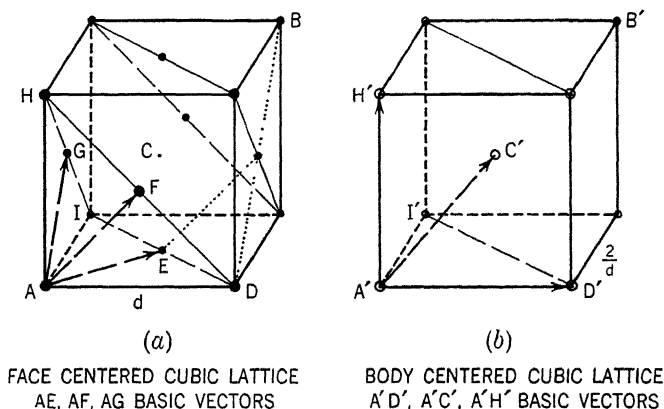


FIG. 32.2.

these lattice planes. The distance between the lattice planes in the direction AD is $d/2$, and hence there is a point D' of the reciprocal lattice in this direction at a distance $2/d$ from the point A' (see Fig. 32.2b). Similarly, there are points at distance $d/2$ from A in the directions AI and AH . The corresponding points are indicated by I' and H' in the reciprocal lattice. There is only one set of planes in the direct lattice separated by a distance greater than $d/2$, viz., those parallel to the plane determined by $DEIGHF$. Thus there can be only one point in the first cube of the reciprocal lattice besides those already found. All other points will be in different cells. The planes are shown in Fig. 32.2a. These planes are separated by a distance

$$\delta = \frac{1}{3} AB = \frac{1}{3} \sqrt{3} d = \frac{d}{\sqrt{3}}$$

and hence

$$\frac{1}{\delta} = \frac{\sqrt{3}}{d} = \frac{\sqrt{3}}{2} \frac{2}{d} = \frac{\sqrt{3}}{2} b \quad \text{where} \quad b = \frac{2}{d}$$

gives the position of a lattice point in the reciprocal lattice in the direction $A'B'$. This point is at the center C' of the cube. This structure—a cube with a point at its center—is called a *body-centered cube*. Thus the reciprocal of the face-centered cubic lattice is a body-centered cubic lattice. It may easily be shown by similar reasoning that the converse statement is true: The reciprocal of a body-centered cubic lattice is a face-centered cubic lattice. The first cell of the body-centered cubic lattice is taken as the parallelepiped determined by the lines $A'D'$, $A'H'$, and $A'C'$. The volume of this cell is one-half the volume of the cube

$$V_b = \frac{1}{2} \left(\frac{2}{d} \right)^3 = \frac{4}{d^3} = \frac{1}{V_d}$$

as it should be. Many other examples may be found in books on crystal structure, for instance, P. P. Ewald's book.

33. Zones in Three Dimensions and Bragg Reflection; Ewald's Construction

The discussion given in Chap. VI can be used without any change for three-dimensional problems. We used vector notation, which works just as well for three as for two components. The areas S_a and S_b are to be replaced by the volumes V_a and V_b as in Eq. (32.5), and atomic planes in the direct lattice replace the atomic rows occurring in two dimensions, as explained in Eq. (32.14).

Triply periodic functions are expanded in triple Fourier series in exactly the same way as in Sec. 26 for two dimensions. The coefficients $C_{h_1 h_2 h_3}$ with three integral indices h_1 , h_2 , h_3 will simply be written C_h for convenience. The proof of the periodicity of ν as a function of \mathbf{a} is carried out exactly as in Sec. 27 and may be briefly repeated here on account of its importance.

A point in the direct lattice is given by

$$\mathbf{R} = l_1 \mathbf{d}_1 + l_2 \mathbf{d}_2 + l_3 \mathbf{d}_3 \quad (33.1)$$

where l_1 , l_2 , and l_3 are integers and \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_3 are the lattice

vectors. An arbitrary vector in the direct lattice is given by

$$\mathbf{r} = \xi_1 \mathbf{d}_1 + \xi_2 \mathbf{d}_2 + \xi_3 \mathbf{d}_3 \quad (33.2)$$

where ξ_1 , ξ_2 , and ξ_3 are the components of \mathbf{r} in the directions \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_3 , respectively. If a wave is propagating through the lattice and can be observed only at the lattice points \mathbf{R} , we may express the disturbance ψ by

$$\psi = A e^{2\pi i [v t - (\mathbf{a} \cdot \mathbf{R})]} \quad (33.3)$$

where \mathbf{a} is the wave vector and is regarded as a vector of the reciprocal lattice. The function ψ is evidently periodic in \mathbf{a} and \mathbf{R} . The vector

$$\mathbf{a}'_h = \mathbf{a} + h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3 \quad (33.4)$$

where h_1 , h_2 , and h_3 are integers, will describe the motion at the lattice points just as well as \mathbf{a} ; for

$$\begin{aligned} (\mathbf{a}'_h \cdot \mathbf{R}) &= (\mathbf{a} \cdot \mathbf{R}) + h_1 (\mathbf{b}_1 \cdot \mathbf{R}) + h_2 (\mathbf{b}_2 \cdot \mathbf{R}) + h_3 (\mathbf{b}_3 \cdot \mathbf{R}) \\ &= (\mathbf{a} \cdot \mathbf{R}) + h_1 l_1 + h_2 l_2 + h_3 l_3 = (\mathbf{a} \cdot \mathbf{R}) + \text{integer} \end{aligned}$$

Since the same motion can be represented by \mathbf{a} or any arbitrary \mathbf{a}'_h , the frequency must be a periodic function of \mathbf{a} also, with the periodicity \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 of the reciprocal lattice.

We wish to set up conventions for eliminating all but one of the values for \mathbf{a} , *i.e.*, define zones to which \mathbf{a} is to be restricted, as was done in Sec. 27 for the two-dimensional case. The method is exactly the same, and we take the first zone to be a volume centered upon the origin O of the reciprocal lattice and limited by plane perpendicular bisectors of vectors in the reciprocal lattice. The first zone has a volume V_b equal to the volume of the elementary cell in the reciprocal lattice, and any point in space can be brought back into the first zone by \mathbf{H} translations in the reciprocal lattice.

$$\mathbf{H} = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3$$

Higher zones surrounding the first one are built in a similar way, and some actual examples will be given below.

The choice of these particular planes, bisecting and perpendicular to the vectors of the reciprocal lattice, is based on the analysis previously given in Secs. 28 and 30 for two dimensions. The theory of waves propagating through a continuous medium

with a small periodic perturbation is identical with the two-dimensional theory, and one must distinguish between

$$\left. \begin{array}{l} \text{Case 1, where all } |\mathbf{a}'_h| \neq |\mathbf{a}| \\ \text{Case 2, where one certain } \mathbf{a}'_h, \text{ say } \mathbf{a}'_n, \text{ is } |\mathbf{a}'_n| \approx |\mathbf{a}| \end{array} \right\} \quad (33.5)$$

where \mathbf{a}'_h is a vector (33.4). Case 2 arises when two vectors \mathbf{a}'_n and \mathbf{a} having almost the same length can be found, such that their difference is a vector of the reciprocal lattice.

$$\mathbf{a}'_n - \mathbf{a} = \mathbf{H} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3 \quad (33.6)$$

A geometrical interpretation of this condition was presented in Sec. 29 where it was proved that the circumstances leading to case 2 were identical with those yielding Bragg's reflections in the crystal. This is always true if atomic rows are replaced by atomic planes in the direct lattice.

Another geometrical interpretation was given by P. P. Ewald.

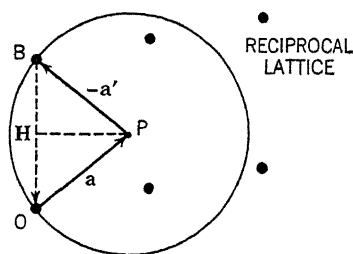


FIG. 33.1.

From point O of the reciprocal lattice he draws a vector \mathbf{a} to point P . A sphere is drawn with radius $|\mathbf{a}|$ about the point P at the end of the vector \mathbf{a} . If it happens that this sphere passes through or near to a second point B of the reciprocal lattice, we obtain a discontinuity in the frequency as a function of \mathbf{a} .

The reason for this is that there is some \mathbf{a}'_n that just fulfills Eq. (33.6). There is no way of telling from the motion of the lattice points which of vectors \mathbf{a} and \mathbf{a}'_n should be preferred. Since the two vectors have different directions, we must assume that the motion is given by a superposition of the two waves. One of these is an incident wave and the other a reflected wave (Bragg wave). There will be two values for the frequency for this single value of $|\mathbf{a}|$ and hence a discontinuity in the ν vs. \mathbf{a} curve.

We note that when Bragg reflection occurs, the vector terminates on a plane that is the perpendicular bisector of the line \mathbf{H} joining two lattice points of the reciprocal lattice. This justifies the rule for constructing zones. We construct planes that are perpendicular bisectors of lines joining lattice points with the

origin. The smallest polyhedron bounded by such planes is the first zone. The area bounded by the first zone and the second smallest polyhedron is the second zone and similarly for higher order zones.

34. General Results for a Wave Propagating in a Three-dimensional Periodic Medium

The problem of wave propagation in a continuous medium with a small periodic perturbation was discussed in Secs. 28 to 30 for two dimensions. The results immediately extend to three dimensions without any difficulty.

We shall now discuss the general problem of wave propagation in a periodic medium, without restricting our discussion to the case of a small periodic perturbation. The following sections will contain various applications and examples of this general theory.

The three-dimensional wave equation is

$$\nabla^2 \psi + \frac{\omega^2}{V^2} \psi = 0 \quad (34.1)$$

We confine ourselves to waves in a periodic medium, and therefore we may assume

$$\frac{1}{V^2} = F(\mathbf{r}) \quad (34.2)$$

where $F(\mathbf{r})$ is periodic in the three directions specified by the lattice vectors \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_3 ; *i.e.*,

$$F(\mathbf{r}) = F(\mathbf{r} + n_1 \mathbf{d}_1 + n_2 \mathbf{d}_2 + n_3 \mathbf{d}_3) \quad n_1, n_2, \text{ and } n_3 \text{ integers}$$

The basis vectors of the reciprocal lattice are \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 , where

$$(\mathbf{b}_i \cdot \mathbf{d}_j) = \delta_{ij}$$

$F(\mathbf{r})$ may be expanded in a triple Fourier sum.

$$F = \sum C_{m_1 m_2 m_3} e^{2\pi i [m_1 (\mathbf{b}_1 \cdot \mathbf{r}) + m_2 (\mathbf{b}_2 \cdot \mathbf{r}) + m_3 (\mathbf{b}_3 \cdot \mathbf{r})]} \quad (34.3)$$

The *solution* of Eq. (34.1) may also be expressed as a Fourier sum.

$$\psi = A(\mathbf{r}) e^{2\pi i [\nu t - (\mathbf{a} \cdot \mathbf{r})]} \quad (34.4)$$

where A is to be periodic in \mathbf{r} and may be written

$$A(\mathbf{r}) = \sum A_{m_1 m_2 m_3} e^{2\pi i [m_1 (\mathbf{b}_1 \cdot \mathbf{r}) + m_2 (\mathbf{b}_2 \cdot \mathbf{r}) + m_3 (\mathbf{b}_3 \cdot \mathbf{r})]} \quad (34.5)$$

The form taken for the ψ solution in Eq. (34.4) will be justified by the following discussion. It is generally known as F. Bloch's theorem and represents the generalization for three dimensions of an old theorem dating back to Floquet for the one-dimensional problem of Mathieu's equation, which is discussed in the next chapter. Combining Eq. (34.4) with Eq. (34.5) yields ψ as a Fourier sum.

$$\psi = e^{i\omega t} \sum A_{m_1 m_2 m_3} e^{-2\pi i(a'_{m_1 m_2 m_3} \cdot r)} \quad (34.6)$$

where

$$\mathbf{a}'_{m_1 m_2 m_3} = \mathbf{a} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2 - m_3 \mathbf{b}_3 \quad (34.7)$$

The last three terms in Eq. (34.7) specify a lattice point in the reciprocal lattice.

Substitution of the solution (34.6) into the wave equation gives a relation between ω and \mathbf{a} . The Laplacian ∇^2 of an arbitrary term of the sum in Eq. (34.6) is

$$\nabla^2 e^{-2\pi i(a'_{m_1 m_2 m_3} \cdot r)} = -4\pi^2 |\mathbf{a}'_{m_1 m_2 m_3}|^2 e^{-2\pi i(a'_{m_1 m_2 m_3} \cdot r)}$$

and substitution in Eq. (34.1) gives

$$\begin{aligned} \nabla^2 \psi &= -4\pi^2 \sum_{m_i} A_{m_1 m_2 m_3} |\mathbf{a}'_{m_1 m_2 m_3}|^2 e^{-2\pi i(a'_{m_1 m_2 m_3} \cdot r)} = -\omega^2 F(\mathbf{r}) \psi \\ &= -\omega^2 \sum_{n_i} C_{n_1 n_2 n_3} e^{2\pi i \Sigma n_i \cdot (b_i \cdot r)} \cdot \sum_{p_i} A_{p_1 p_2 p_3} e^{-2\pi i[(a \cdot r) - \Sigma p_i \cdot (b_i \cdot r)]} \end{aligned} \quad (34.8)$$

replacing $\mathbf{a}'_{p_1 p_2 p_3}$ by its equivalent given in Eq. (34.7). We introduce new subscripts m_1 , m_2 , and m_3 in the last form of Eq. (34.8) as follows:

$$m_i = n_i + p_i \quad (34.9)$$

Equation (34.8) may now be written in the form

$$\begin{aligned} \frac{4\pi^2}{\omega^2} \sum_{m_i} A_{m_1 m_2 m_3} |\mathbf{a}'_{m_1 m_2 m_3}|^2 e^{-2\pi i(a'_{m_1 m_2 m_3} \cdot r)} \\ = \sum_{m_i} \sum_{p_i} C_{m_1 - p_1, m_2 - p_2, m_3 - p_3} A_{p_1 p_2 p_3} e^{-2\pi i(a'_{m_1 m_2 m_3} \cdot r)} \end{aligned} \quad (34.10)$$

and equating corresponding terms, we obtain, finally,

$$\frac{1}{\nu^2} A_m |\mathbf{a}'_m|^2 = \sum_{p=-\infty}^{\infty} C_{m-p} A_p \quad (34.11)$$

where one should remember that expansion (34.3) must represent a *real* function F , hence the relation

$$C_{-n} = C_n^*$$

The equation is obviously *symmetrical in all* $|\mathbf{a}'_m|^2$, and so are the solutions

$$\nu^2 = f(\cdots |\mathbf{a}'_m|^2 \cdots) \quad (34.16)$$

Some general results are thus provided:

a. The frequency ν is a periodic function of the \mathbf{a} vectors, with the periods \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 of the reciprocal lattice, since replacing \mathbf{a} by any \mathbf{a}'_m just changes the names of the \mathbf{a} vectors without changing the set of \mathbf{a} vectors as a whole. This is also shown by the fact that the wave itself [Eq. (34.6)] contains all the \mathbf{a}'_m vectors in a similar way and offers no possibility of assigning to any one of these vectors special importance.

b. Changing \mathbf{a} into $-\mathbf{a}$ makes a similar change in all vectors, since

$$-\mathbf{a}'_m = -(\mathbf{a} - m_1\mathbf{b}_1 - m_2\mathbf{b}_2 - m_3\mathbf{b}_3) = -\mathbf{a} + m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 = (-\mathbf{a})'_{-m} \quad (34.17)$$

Hence, the same frequency is obtained for two similar waves propagating in opposite directions.

c. The frequency ν depends *symmetrically* upon all vectors \mathbf{a}'_m and contains *only the absolute value of these vectors*, not their direction. This supports the rule used for the definition of the boundary of the zones, which was based only on the existence of certain relations between the lengths of the vectors \mathbf{a}'_m . The zone boundaries are obtained for a small periodic perturbation, as discussed in the next section.

$$\text{Zone boundary:} \quad |\mathbf{a}| = |\mathbf{a}'_m| \quad (34.18)$$

The definition of the first zone, for instance, is justified in this way. Among all the \mathbf{a}'_m entering the expansion of the wave [Eq. (34.6)] we chose the smallest $|\mathbf{a}|$ as the parameter. This special vector \mathbf{a} is used to specify the wave, but for each \mathbf{a} we obtain an infinite number of waves, with an infinite number of frequencies that are the roots of the infinite determinant [Eq. (34.14)].

These infinite determinants will be encountered in the theory of Mathieu's equation, which will be discussed in the next

chapter. The Mathieu problem corresponds to one dimension when the determinant (34.15) is obtained, and this special case has been completely computed by Hill and Whittaker (see Sec. 43). It would be very interesting to obtain an extension of Hill's results for three dimensions, but this problem has not yet been discussed.

We shall now examine the problem for the case of a small periodic perturbation.

35. Waves in a Homogeneous Isotropic Medium with a Small Periodic Perturbation

The problem of the propagation of waves through a three-dimensional homogeneous isotropic medium with a small periodic perturbation does not differ from that for two dimensions treated in Sec. 28. We assume that $1/V^2$ is almost constant with a small periodic variation.

$$\frac{1}{V^2} = \frac{1}{V_0^2} + \epsilon f \quad (35.1)$$

To continue the discussion we expand f in Eq. (35.1) as a Fourier sum.

$$f = \sum c_{m_1 m_2 m_3} e^{2\pi i \sum m_i (b_i \cdot r)} \text{ with } c_{000} = 0 \quad (35.2)$$

Then $C_{n_1 n_2 n_3}$ in the expansion of $F(\mathbf{r}) = 1/V^2$ may be written in the following form:

$$C_{n_1 n_2 n_3} = \frac{1}{V_0^2} \delta_{0n_1} \delta_{0n_2} \delta_{0n_3} + \epsilon c_{n_1 n_2 n_3} \quad (35.3)$$

with Kronecker symbols δ , which mean

$$\left. \begin{aligned} C_{000} &= \frac{1}{V_0^2} && \text{finite} \\ \text{and all other } C_{n_1 n_2 n_3} &= \epsilon c_{n_1 n_2 n_3} && \text{infinitely small} \end{aligned} \right\} \quad (35.4)$$

This means that the determinant (34.14) or (34.15) has finite terms only along the principal diagonal, while all off-diagonal terms are infinitely small. We want to expand the determinant in powers of ϵ . The term independent of ϵ is the diagonal term

$$\prod_m \left(\frac{C_{000}}{|\mathbf{a}'_m|^2} - \frac{1}{v^2} \right) \quad (35.5)$$

which we may write as

$$\prod_m \left(\frac{1}{\nu_m^2} - \frac{1}{\nu^2} \right); \quad \frac{1}{\nu_m^2} = \frac{C_{000}}{|\mathbf{a}'_m|^2} = \frac{1}{V_0^2 |\mathbf{a}'_m|^2} \quad (35.6)$$

The next terms in the expansion are in ϵ^2 and are obtained by replacing the terms n and p on the diagonal by the $(n \cdot p)$ and $(p \cdot n)$ nondiagonal terms. These terms have a minus sign according to the law of determinants.

$$\begin{aligned} -\epsilon^2 \sum_{n,p} \prod_{\substack{m \neq n \\ m \neq p}} \left(\frac{1}{\nu_m^2} - \frac{1}{\nu^2} \right) \frac{c_{n-p} c_{p-n}}{|\mathbf{a}'_n|^2 |\mathbf{a}'_p|^2} \\ = -\epsilon^2 V_0^4 \sum_{n,p} \prod_{\substack{m \\ n \neq m \neq p}} \left(\frac{1}{\nu_m^2} - \frac{1}{\nu^2} \right) \frac{|c_{n-p}|^2}{\nu_n^2 \nu_p^2}. \end{aligned} \quad (35.7)$$

To make this process clear, let us consider the determinant (34.15) and write the term $n = 0, p = 1$.

$$\dots \left(\frac{C_0}{|\mathbf{a}'_{-2}|^2} - \frac{1}{\nu^2} \right) \left(\frac{C_0}{|\mathbf{a}'_{-1}|^2} - \frac{1}{\nu^2} \right) \left(-\epsilon^2 \frac{c_{-1}}{|\mathbf{a}'_0|^2} \frac{c_1}{|\mathbf{a}'_1|^2} \right) \left(\frac{C_0}{|\mathbf{a}'_2|^2} - \frac{1}{\nu^2} \right) \dots$$

We shall not attempt to compute terms in $\epsilon^3, \epsilon^4, \dots$. Grouping terms (35.6) and (35.7) together, we obtain for our determinant

$$\begin{aligned} D = 0 = \sum_{n,p} \prod_{\substack{m \\ n \neq m \neq p}} \left(\frac{1}{\nu_m^2} - \frac{1}{\nu^2} \right) \left[\left(\frac{1}{\nu_n^2} - \frac{1}{\nu^2} \right) \left(\frac{1}{\nu_p^2} - \frac{1}{\nu^2} \right) - \epsilon^2 V_0^4 \frac{|c_{n-p}|^2}{\nu_n^2 \nu_p^2} \right] \end{aligned} \quad (35.8)$$

We can now discuss the solution of this equation. The *first approximation* is obviously obtained by making the product (35.6) of the diagonal terms zero, which means that ν must be equal to one of the ν_m 's:

$$\nu = \nu_m \quad (35.9)$$

This is correct as long as no two values ν_n and ν_p are too close together, which means

Case 1:

$$\text{All } \nu_n \neq \nu_p \quad \text{or all } |\mathbf{a}'_n|^2 \neq |\mathbf{a}'_p|^2 \quad (35.10)$$

An exception arises when two values ν_n and ν_p are close enough to make both terms of equal order of magnitude in one of the square brackets of expansion (35.8).

Case 2:

$$\nu_n \approx \nu_p \quad \text{or} \quad |\mathbf{a}'_n|^2 \approx |\mathbf{a}'_p|^2 \quad (35.11)$$

When this occurs, we can no longer neglect the ϵ^2 terms, and instead of simply writing

$$\left(\frac{1}{\nu_n^2} - \frac{1}{\nu^2}\right) \left(\frac{1}{\nu_p^2} - \frac{1}{\nu^2}\right) = 0$$

we must take the complete square bracket from Eq. (35.8) and write

$$\left(\frac{1}{\nu_n^2} - \frac{1}{\nu^2}\right) \left(\frac{1}{\nu_p^2} - \frac{1}{\nu^2}\right) - \epsilon^2 V_0^4 \frac{|c_{n-p}|^2}{\nu_n^2 \nu_p^2} = 0 \quad (35.12)$$

an equation that includes both case 2 of Sec. 28 and the case discussed in Sec. 30.

In order to see this clearly, let us use the same notation as in Chap. VI, Secs. 28 and 30, and choose one of our two vectors, \mathbf{a}'_n for instance, as vector \mathbf{a} . This is always allowed on account of the equivalence of all vectors \mathbf{a}'_m . Hence we write, instead of (35.11),

$$\nu_n = \nu_0, \quad |\mathbf{a}'_n| = |\mathbf{a} - n_1 \mathbf{b}_1 - n_2 \mathbf{b}_2 - n_3 \mathbf{b}_3| \approx |\mathbf{a}| \quad (35.13)$$

To make the comparison with Sec. 30 easier, let us call \mathbf{a}_0 the vector for which the relation (35.13) would be exactly satisfied and take

$$|\mathbf{a}_0 - \mathbf{B}| = |\mathbf{a}_0|, \quad \mathbf{B} = n_1 \mathbf{b}_1 + n_2 \mathbf{b}_2 + n_3 \mathbf{b}_3 \quad (35.14)$$

Then the case of the approximate condition (35.13) is defined by

$$\left. \begin{aligned} \mathbf{a} &= \mathbf{a}_0 + \eta \mathbf{B} \\ \mathbf{a}'_n &= \mathbf{a}_0 - \mathbf{B} + \eta \mathbf{B} \end{aligned} \right\} \quad \eta \text{ small} \quad (35.15)$$

which corresponds to the definitions (30.1) and yields, as in Eqs. (30.3) and (30.4),

$$\left. \begin{aligned} |\mathbf{a}|^2 &= |\mathbf{a}_0|^2 + 2\eta(\mathbf{a}_0 \cdot \mathbf{B}) \\ |\mathbf{a}'_n|^2 &= |\mathbf{a}_0|^2 - 2\eta(\mathbf{a}_0 \cdot \mathbf{B}) \end{aligned} \right\} \quad (35.16)$$

or

$$\left. \begin{aligned} \nu^2 &= \nu_0^2 + 2\eta V_0^2(\mathbf{a}_0 \cdot \mathbf{B}) \\ \nu_n^2 &= \nu_0^2 - 2\eta V_0^2(\mathbf{a}_0 \cdot \mathbf{B}) \end{aligned} \right\} \quad (35.17)$$

With this notation, Eq. (35.12) can be transcribed as

$$\left[\frac{1}{\nu_0^2} + \frac{2\eta}{\nu_0^4} V_0^2(\mathbf{a}_0 \cdot \mathbf{B}) - \frac{1}{\nu^2} \right] \left[\frac{1}{\nu_0^2} - \frac{2\eta}{\nu_0^4} V_0^2(\mathbf{a}_0 \cdot \mathbf{B}) - \frac{1}{\nu^2} \right] - \epsilon^2 V_0^4 \frac{|c_n|^2}{\nu_n^2 \nu^2} = 0$$

but $\nu_n^2 \nu^2$ can be replaced by ν_0^4 in the last term, which is already small like ϵ^2 .

$$\left(\frac{1}{\nu^2} - \frac{1}{\nu_0^2} \right)^2 - \frac{4\eta^2}{\nu_0^8} V_0^4 (\mathbf{a}_0 \cdot \mathbf{B})^2 - \epsilon^2 V_0^4 \frac{|c_n|^2}{\nu_0^4} = 0 \quad (35.18)$$

which yields

$$\frac{1}{\nu^2} = \frac{1}{\nu_0^2} \pm \epsilon \frac{V_0^2}{\nu_0^4} \sqrt{\nu_0^4 |c_n|^2 + [2h(\mathbf{a}_0 \cdot \mathbf{B})]^2}, \quad h = \frac{\eta}{\epsilon} \quad (35.19)$$

This is exactly the same as Eq. (30.13), where we had

$$\omega^2 = \omega_0^2 + \epsilon k_1 \quad (28.7)$$

$$k_1 = \pm V_0^2 4\pi^2 \sqrt{\nu_0^4 |c_n|^2 + [2h(\mathbf{a}_0 \cdot \mathbf{B})]^2} \quad (30.13)$$

which means

$$\nu^2 = \nu_0^2 + \frac{\epsilon}{4\pi^2} k_1$$

$$\frac{1}{\nu^2} = \frac{1}{\nu_0^2} - \frac{\epsilon}{4\pi^2 \nu_0^4} k_1$$

and checks with Eq. (35.19).

Thus the general method developed in Sec. 34 checks completely with the approximation previously discussed in Secs. 28 and 30. Just as in this previous discussion, case 1 is characterized by one amplitude coefficient A much larger than all other A coefficients in expansion (35.6), and the definition that gives the best connection with the problem of the uniform homogeneous medium consists in calling A_{000} this specially large coefficient, as was done in Sec. 28. In case 2, two A coefficients are of the same order of magnitude, and according to the notation (35.14) one of them is called A_{000} and the other one $A_{n_1 n_2 n_3}$.

The whole theory of zone structure was based on the discussion of Secs. 28 and 30 and finds here a more complete and general foundation.

36. General Remarks on Waves in a Discontinuous Lattice

We have defined in Sec. 31 a *lattice with basis* for two dimensions (Fig. 31.8), and the same definition applies for three dimensions. The whole lattice is based on three vectors \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_3 as before, and inside each \mathbf{d} cell there is a certain number N of particles, whose positions with respect to the origin of the cell are

$$\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$$

and whose masses are taken as M_1, M_2, \dots, M_N , respectively. If $N = 1$, we obtain the monatomic lattice, with one particle per cell, and it is convenient to take $\mathbf{r}_1 = 0$ and have this particle at the origin of the cell. In the general case, an arbitrary particle k anywhere in the lattice is located at

$$\mathbf{r} = \mathbf{r}_j + n_1\mathbf{d}_1 + n_2\mathbf{d}_2 + n_3\mathbf{d}_3 \quad (36.1)$$

where particle k is the j th particle in cell (n_1, n_2, n_3) . A general scheme for propagation of waves through such a lattice was given in Sec. 17 for the one-dimensional case. It could be developed in a similar way for three dimensions. For each j type of particle in the lattice we could write a wave

$$\psi_j = A'_j e^{2\pi i(\nu t - \mathbf{a} \cdot \mathbf{r})} \quad (36.2)$$

but it is more convenient to include the exponential $e^{-2\pi i(\mathbf{a} \cdot \mathbf{r}_j)}$ in the complex amplitude A_j and to keep the same imaginary exponential in all the j waves.

$$\left. \begin{aligned} & e^{i(\omega t - n_1 k_1 - n_2 k_2 - n_3 k_3)} \\ k_1 = 2\pi(\mathbf{a} \cdot \mathbf{d}_1), \quad k_2 = 2\pi(\mathbf{a} \cdot \mathbf{d}_2), \quad k_3 = 2\pi(\mathbf{a} \cdot \mathbf{d}_3) \end{aligned} \right\} \quad (36.3)$$

as we did in Sec. 27 for two dimensions. The wave motion of particles j is then described by

$$\psi_j = A_j e^{i(\omega t - n_1 k_1 - n_2 k_2 - n_3 k_3)}, \quad A_j = A'_j e^{-2\pi i(\mathbf{a} \cdot \mathbf{r}_j)} \quad (36.4)$$

There will be N constants: A_1, A_2, \dots, A_N , in general complex to include the phase of particle j with particle 1.

As in the one-dimensional case we will have N equations of motion, and substitution of Eq. (36.4) in the equations of motion will yield N linear equations in the A_i . Each A_i is a vector function and has therefore three components, thus yielding altogether $3N$ linear homogeneous equations to be solved simul-

taneously. As before, we equate the determinant of the coefficients of the A_i to zero, which yields an equation of degree $3N$

in ω^2 . This means that for each value of a in the first zone of the reciprocal lattice, there will be $3N$ values for the frequency. This is shown schematically in Fig. 36.1.

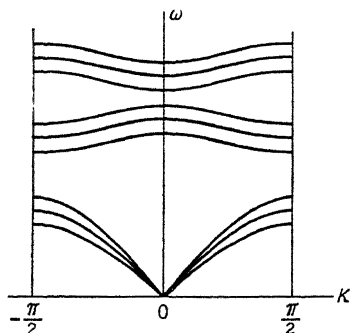


FIG. 36.1.

The number of branches for the function $\nu(\mathbf{a})$ is thus equal to three times the number N of particles per cell. A continuous periodic medium may be regarded as the limit $N \rightarrow \infty$. Uncertainty in the wave vector \mathbf{a} is best avoided by restricting \mathbf{a} to the interior of the first zone.

37. Some Examples of Zones in Three Dimensions

The face-centered and body-centered cubic lattices were discussed in Sec. 32 (Fig. 32.2) where it was proved that they are mutually reciprocal. The *face-centered cubic lattice* is of special

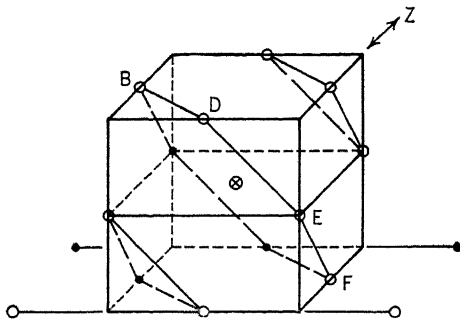


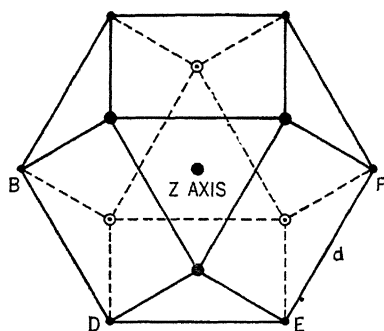
FIG. 37.1.—Face-centered cubic lattice.

importance since it represents one of the possible structures for close-packed spheres. The lattice can be described as in Fig. 32.2a, and also as in Fig. 37.1, which obviously represents the same structure since it is derived from Fig. 32.2a by translation of $\frac{1}{2}d$ parallel to the axis. Looking at this cubic structure from the direction of the arrow and taking the diagonal OZ as the vertical axis, we obtain the drawing of Fig. 37.2, which

represents the same lattice as the hexagonal structure shown in Fig. 37.3. This is one of the possible structures obtained by piling up equal spheres provided that the distance from the central sphere to its *twelve neighbors* is the same, and this is the case for

$$\frac{c}{d} = 2\sqrt{\frac{2}{3}} = 1.633 \quad (37.1)$$

where the distance c measured vertically from A to B is slightly less than the distance $BE = \sqrt{3}d = 1.732d$. The structure repeats itself after a vertical distance $(\frac{3}{2})c$.



- • • Main Horizontal Plane
- • • Top
- ⊙ ⊙ ⊙ Bottom

FIG. 37.2.—Face-centered cubic lattice.

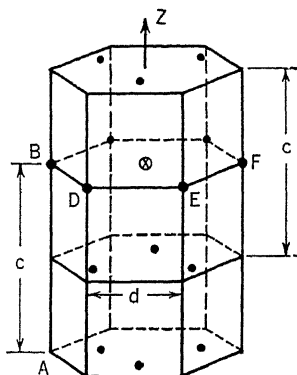


FIG. 37.3.—Face-centered cubic lattice.

There is *another possible structure* for *close-packed spheres* shown in Figs. 37.4 and 37.5. It obviously has exactly the same density as the face-centered cubic lattice. Comparing Fig. 37.3 with Fig. 37.4, we notice that in Fig. 37.3 the lower atoms are diagonally opposite to the upper ones. In Fig. 37.4, on the other hand, the lower atoms lie just below the upper ones, and the structure repeats itself after a vertical distance c . This remark shows immediately that the *hexagonal lattice* of Figs. 37.4 and 37.5 is *not a simple Bravais lattice* built on three vectors \mathbf{d}_1 , \mathbf{d}_2 , and \mathbf{d}_3 , but that it represents a *lattice with basis* according to the definition given in Sec. 31 (Fig. 31.8).

In the case of lattices containing only one type of particle, the following criterion can be used to distinguish between a simple

Bravais lattice and lattices with a basis: Taking a vector OA from one particle O to another A , we construct its negative OA' (Fig. 37.6). If the lattice is a simple Bravais lattice, there must be an atom at the point A' (as in Figs. 37.2 and 37.3). If there is no atom at A' (as in Figs. 37.4 and 37.5), the lattice has a basis. Hence, the face-centered cubic structure, which

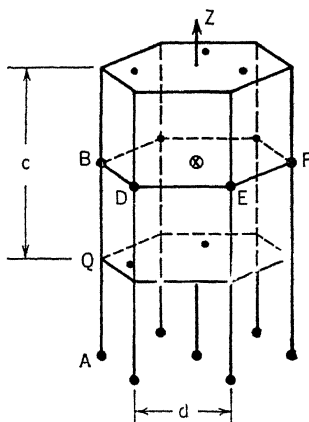
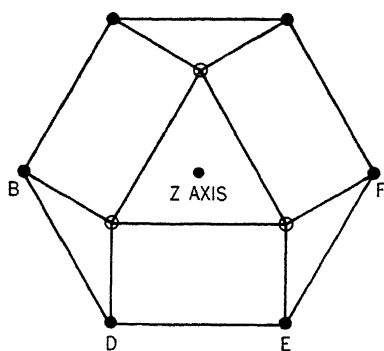


FIG. 37.4.—Hexagonal close-packed lattice.



● ● ● Main Horizontal Plane
○ ○ ○ Top and Bottom

FIG. 37.5.—Hexagonal close-packed lattice.

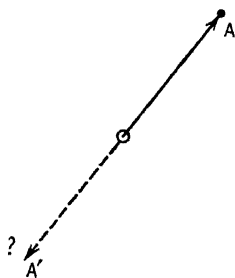


FIG. 37.6.

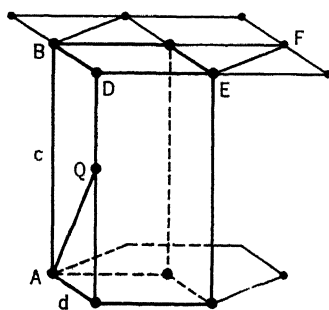


FIG. 37.7.—Hexagonal close-packed lattice.

looked like a structure with a basis, actually represents a simple Bravais lattice. The hexagonal close-packed lattice, however, cannot be represented that way and actually possesses a basis. The fundamental cell contains two atoms as shown in Fig. 37.7 where AQ represents the basis vector. In this hexagonal lattice as in the face-centered cubic lattice of Fig. 37.3

$$AB = c = 1.633d, \quad BE = 1.732d$$

for spherical particles. Both structures can be projected on a vertical plane passing through DF as shown in Fig. 37.8. In the reciprocal-lattice discussion of a lattice with a basis, all additional atoms inside the basis must be ignored, and the discussion refers only to the three fundamental vectors d_1 , d_2 , and d_3 . In the hexagonal structure of Fig. 37.7, for instance, we ignore the Q points on the level $\frac{1}{2}c$ and look only for lattice planes running through the points on levels O or c . In a horizontal projection,

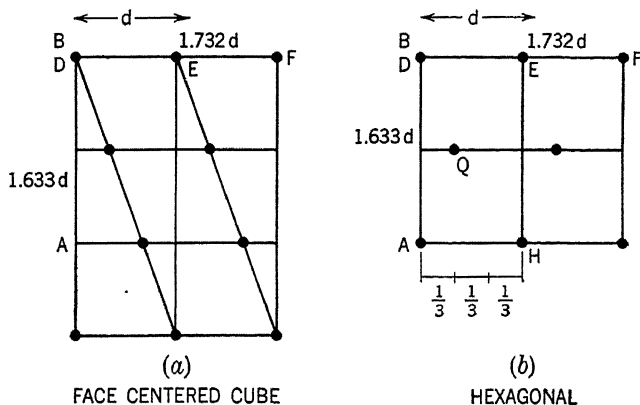


FIG. 37.8.

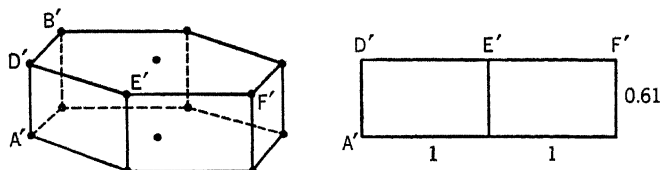


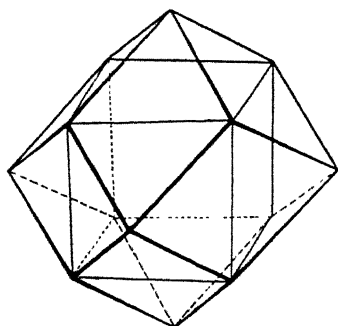
FIG. 37.9.—Hexagonal close-packed lattice.

the hexagonal structure is the same as in Fig. 31.2 for two dimensions and the reciprocal is another hexagonal structure with the side

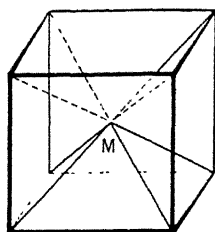
$$b = \frac{2}{\sqrt{3}d} \quad (37.2)$$

as drawn in Fig. 31.3. Looking at the vertical projection in Fig. 37.8b, we obtain the fundamental cell represented by $ABEH$ with dimensions

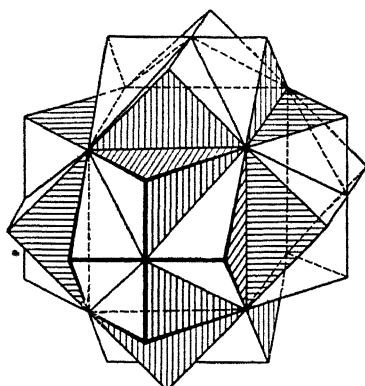
$$BE = \frac{\sqrt{3}}{2}d \quad \text{and} \quad AB = c = 2\sqrt{\frac{2}{3}}d \quad (37.3)$$



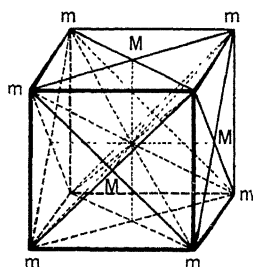
1st zone—central cube
2nd zone—between central cube and dodecaeder



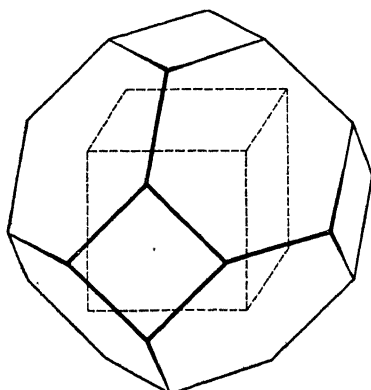
Reduction of 2nd zone



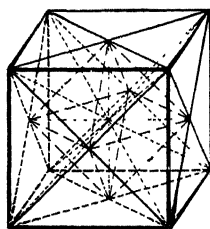
3rd zone—between dodecaeder and outer surface



Reduction of 3rd zone



4th zone—between 3rd zone and new outer surface



Reduction of 4th zone

FIG. 37.10.—Cubic lattice, zones 1, 2, 3 and 4.

This gives us the reciprocal lattice shown in Fig. 37.9, a flat hexagonal structure whose vertical projection is also given.

Let us now discuss the *zone structures*. Starting with a *cubic lattice*, we obtain the zones represented in Fig. 37.10, where the first four surfaces have been drawn and the reduction to the first cubic zone is indicated on the right with the three-dimensional mosaics obtained by translations parallel to the side of the cube.

A *face-centered cubic lattice* has a body-centered cubic lattice as its reciprocal. The first zone is shown in Fig. 37.11 and has the same shape as the fourth zone of the cubic lattice. The fundamental cell of the direct lattice has a volume $\frac{1}{4}d^3$ (four

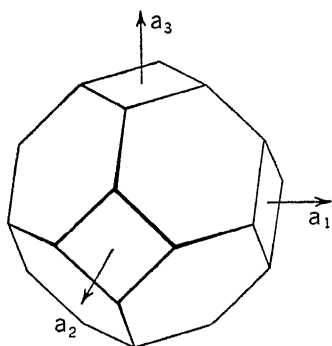


FIG. 37.11.—Face-centered cubic lattice, first zone.

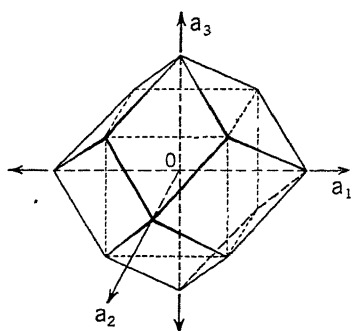


FIG. 37.12.—Body-centered cubic lattice, first zone.

points per cube); hence the first zone and each fundamental cell of the reciprocal lattice have a volume of $4/d^3$.

A *body-centered cubic lattice* has a *face-centered cubic lattice* as its reciprocal. The first zone is shown in Fig. 37.12 and is similar to the second zone of the cubic lattice. The volumes are $\frac{1}{2}d^3$ for the direct lattice and $2/d^3$ for the reciprocal lattice and the first zone. The second zone is similar to Fig. 37.11.

A hexagonal lattice with arbitrary d and c has a hexagonal reciprocal with $2/\sqrt{3}d$ and $1/c$ as shown in Fig. 37.9. The first zone is a hexagonal cell looking exactly like Fig. 37.9. First and second zones are shown in Fig. 37.13.

Let us now consider the transition from a continuous medium to a discontinuous lattice. Starting from a homogeneous isotropic continuum and a certain type of wave (longitudinal elastic waves, for instance), we first introduce a slight periodic

perturbation with the face-centered cubic distribution. This introduces discontinuities in the $\nu(\mathbf{a})$ relation on all plane surfaces corresponding to zone boundaries. Instead of letting \mathbf{a} run from 0 to ∞ , we may reduce all vectors \mathbf{a} inside the first zone, since each higher zone possesses sections that give a continuous $\nu(\mathbf{a})$ branch inside the first zone, and finally we obtain an infinite set of successive branches inside the first zone. This is the three-dimensional equivalent of the process described by Fig. 17.2 for one dimension. Increasing the periodic perturbation, we finally reduce the structure to the discrete atoms located at the points of the face-centered lattice. In this process all the upper branches of the $\nu(\mathbf{a})$ curves rise to infinity and disappear. The only remaining branch is the lower one, and

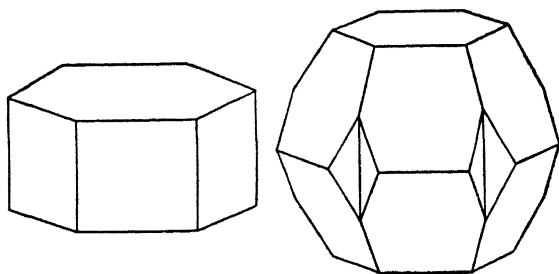


FIG. 37.13.—Hexagonal lattice, first and second zone.

Fig. 17.2 reduces to Fig. 2.4*b*. A very similar process of transformation was discussed at the beginning of Sec. 16.

Let us now follow the same procedure with a hexagonal close packing. The periodicity is hexagonal, but there are two atoms in each cell. The transformation will leave us with *two branches* in the first zone instead of one. The final $\nu(\mathbf{a})$ will be similar to the double curve or the one-dimensional NaCl lattice (Fig. 16.1), instead of the single curve of Fig. 2.4*b*. In Sec. 16 we discussed the transition from the case of two different masses $M_1 \neq M_2$ to two equal masses $M_1 = M_2$, and Fig. 16.1 explained the transformation. A similar discussion could be applied to the transformation indicated in Fig. 37.14 and obtained by straightening the vertical atomic line, thus going from the close-packed hexagonal lattice of height c to the simple Bravais hexagonal lattice of height $\frac{1}{2}c$. The fundamental cell of the latter has a volume one-half that of the original. Its first zone has a double

volume, being twice as high as the original one of Fig. 37.13. In this first zone we should find just *one* $\nu(\mathbf{a})$ branch, but as soon as the perturbation corresponding to the zigzag of the vertical lines is introduced, this zone is cut in two, and the single branch splits into two branches in the new first zone. The procedure

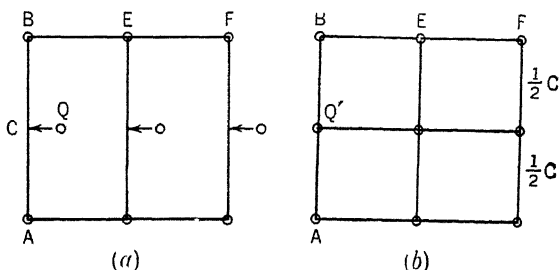


FIG. 37.14.

along the z axis is exactly the same as in Born's discussion of the NaCl reduction.

38. Zones in the Direct Lattice; Principle of the Wigner-Seitz Method

Let us return to the problem of wave propagation in a periodic medium for three dimensions, as discussed in Sec. 34. We found that the general solution could be written as in Eq. (34.4).

$$\psi = A(\mathbf{r})e^{2\pi i[\nu t - (\mathbf{a} \cdot \mathbf{r})]} \quad (34.4)$$

with an amplitude $A(\mathbf{r})$ that is a periodic function in the direct lattice. This offers the possibility of defining the function $A(\mathbf{r})$ inside a single d cell or inside a closed surface containing a volume equal to that of an elementary d cell.

This is the basis of the Wigner-Seitz method of treating the theory of solids that we shall discuss briefly in this section. To do this, we break the vector \mathbf{r} of a point in the cell $(n_1 n_2 n_3)$ into a number of terms.

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{r}' + n_1 \mathbf{d}_1 + n_2 \mathbf{d}_2 + n_3 \mathbf{d}_3 \quad (38.1)$$

\mathbf{r}_0 is the vector for some point in the first cell relative to the origin of the cell, $n_1 \mathbf{d}_1 + n_2 \mathbf{d}_2 + n_3 \mathbf{d}_3$ the vector for the origin of the cell $(n_1 n_2 n_3)$ relative to the origin of the lattice, and \mathbf{r}' the coordinate of the point under consideration relative to the point \mathbf{r}_0 in the cell $(n_1 n_2 n_3)$. \mathbf{r}' is to be restricted to values inside

a single cell or an equivalent surface. This is illustrated in Fig. 38.1.

Now we must rewrite the solution (34.4) in terms of \mathbf{r}' . First, we note that $(\mathbf{a} \cdot \mathbf{r})$ becomes

$$(\mathbf{a} \cdot \mathbf{r}) = (\mathbf{a} \cdot \mathbf{r}_0) + (\mathbf{a} \cdot \mathbf{r}') + n_1(\mathbf{a} \cdot \mathbf{d}_1) + n_2(\mathbf{a} \cdot \mathbf{d}_2) + n_3(\mathbf{a} \cdot \mathbf{d}_3) \quad (38.2)$$

We set

$$2\pi(\mathbf{a} \cdot \mathbf{d}_i) = k_i \quad (38.3)$$

Then

$$A(\mathbf{r}) = A(\mathbf{r}_0 + \mathbf{r}') = e^{2\pi i[(\mathbf{a} \cdot \mathbf{r}_0) + (\mathbf{a} \cdot \mathbf{r}')] } A''(\mathbf{r}') \quad (38.4)$$

and we regard $A''(\mathbf{r}')$ as defined in the first cell only. Substituting the expression $A(\mathbf{r})$ into Eq. (34.4), we obtain

$$\psi = A''(\mathbf{r}') e^{i(\omega t - n_1 k_1 - n_2 k_2 - n_3 k_3)} \quad (38.5)$$

This is the wave function inside the cell (n_1, n_2, n_3) . The amplitude A'' is defined inside the first cell and is independent of (n_1, n_2, n_3) .

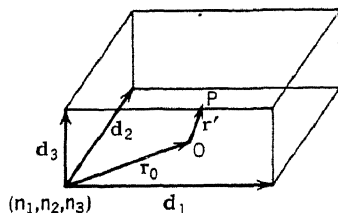


FIG. 38.1.

One may notice immediately that Eq. (38.5) is the translation for a continuous medium of the type of definitions recommended in Sec. 36 for a discontinuous structure. The choice of the first cell is, of course, arbitrary since the choice of the basis vectors

in the direct lattice is arbitrary. Wigner and Seitz choose their elementary cell in a manner similar to that in which zones are constructed for the reciprocal lattice, *viz.*, by constructing a polyhedron about a lattice point. The faces of the polyhedron are planes that are perpendicular bisectors of the lines joining the point taken as the origin with neighboring lattice points. This is shown schematically in Fig. 38.2 for a two-dimensional lattice.

Some conditions are required to insure continuity of the ψ function throughout the lattice and to prevent any discontinuity across the border of the zone in the direct lattice. Let \mathbf{r}' be a point on the boundary; the vector \mathbf{r} defining the analogous point in the next cell (Fig. 38.2) is then

$$\mathbf{r} = \mathbf{r}' + n_1 \mathbf{d}_1 + n_2 \mathbf{d}_2 + n_3 \mathbf{d}_3 \quad (38.6)$$

This results from the definition of the zones in the direct lattice, and n_1 , n_2 , and n_3 are small integers such as 0 or ± 1 if the first zone is being used. We must obtain a continuity condition for ψ and its normal derivative at \mathbf{r}' . The condition

$$A''(\mathbf{r}') = A''(\mathbf{r})e^{-i(n_1k_1+n_2k_2+n_3k_3)} \quad (38.7)$$

ensures the continuity of the wave function ψ . A similar condition holds for the normal derivative. For very long wave length, k_1 , k_2 , and k_3 are small, as are n_1 , n_2 , and n_3 , and one may *approximately* state

$$A''(\mathbf{r}') = A''(\mathbf{r}) \quad (38.8)$$

along the boundary.

This is a first approximation, but it is generally necessary to use a second simplifying assumption, since the polyhedron

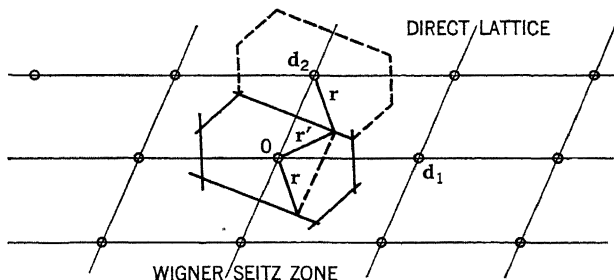


FIG. 38.2.

limiting the Wigner-Seitz zone offers considerable difficulty as a boundary. The simplification used is to replace the polyhedron by a *sphere* of equal inside volume and radius R . The corresponding points \mathbf{r}' and \mathbf{r} are replaced by diametrically opposite points $\pm \mathbf{r}'$.

$$\left. \begin{aligned} A''(\mathbf{r}') &= A''(-\mathbf{r}') \\ \frac{\partial A''}{\partial R} &= 0 \end{aligned} \right\} \quad (38.9)$$

This can give only a first approximation for long wave lengths, and the discussion of corrections for small wave lengths offers serious difficulties.

39. Frequency Distribution for Waves in an Actual Crystal

The general discussion of the preceding sections leads to very important results, which play a prominent role in a number of

problems in theoretical physics. We do not intend to discuss all these problems, but we shall select the oldest one as the most typical: the theory of specific heat of solid bodies. Let us first summarize our main results:

For any kind of waves (elastic, electromagnetic, or wave mechanical De Broglie waves) propagating in a medium with periodic structure, discontinuities are found in the relation between the frequency ν and the wave vector \mathbf{a} (a vector of length $|\mathbf{a}| = 1/\lambda$ and pointing in the direction of propagation). These discontinuities are obtained whenever the vector \mathbf{a} terminates on a plane that is a perpendicular bisector of one of the vectors of the reciprocal lattice. These planes play the same role for all waves, whatever their particular nature may be. For instance, in electromagnetic waves the discontinuities in $\nu(\mathbf{a})$ are directly responsible for the selective reflection of X rays (Bragg-Laue spots). This general property is the reason for choosing these special planes as limits of the *zones* defined in the preceding section, and the zone structure in a given crystal lattice is the same for all waves.

Elastic vibrations, for instance, are propagated through a crystal lattice as *elastic waves*. Their properties result from the discussions of Chaps. II, III, and IV and their generalizations in three dimensions. Assuming N atoms per lattice cell (lattice with basis), the best representation of the waves is obtained by restricting the \mathbf{a} vector to the *first zone*. Each \mathbf{a} vector yields $3N$ different wave motions, of which three are of the acoustic type and $3(N - 1)$ of the optical type, as Fig. 36.1 shows schematically for the case $N = 3$. The three acoustic waves correspond to the well-known waves in an isotropic solid: one longitudinal and two transverse vibrations. In an ideal continuous anisotropic medium three vibrations at right angles are obtained, none of which is exactly longitudinal or transverse. In the crystal lattice with discontinuous structure, there are still three different acoustic waves for each \mathbf{a} vector, but their properties are much more complicated than for a continuous medium. We may, for the sake of visualization, call the waves longitudinal (l) and transverse (t_1, t_2), but these names do not correspond exactly to the properties of waves in a lattice structure.

The hexagonal lattice, although it contains only one type of atom, is a lattice with basis and yields three acoustic branches

and three branches of higher frequency, which may be called *optical*. A real *monatomic* crystal lattice (such as the face- or body-centered cubic structure) is one without basis and contains only one atom per cell. It yields only three acoustic branches and no optical branches.

The volume of the first zone equals the volume V_b of the fundamental cell in the reciprocal lattice, and

$$V_b = \frac{1}{V_d} \quad (39.1)$$

where V_d is the volume of the cell in the direct lattice.

All this refers to an infinite crystal lattice. Now what can be said about a finite piece of crystal of volume V , containing, for instance, one gram molecule and a total of \mathfrak{N} atoms?

$$\mathfrak{N} = 6.06 \times 10^{23} \quad \text{Avogadro's number} \quad (39.2)$$

The vibrations of such a bounded crystal lattice depend upon the properties of the infinite lattice and upon the *boundary conditions*. These conditions are usually very troublesome in all problems of elasticity. Even in the case of an isotropic continuum, boundary conditions generally result in mixing all the wave types. A longitudinal wave, for instance, falling upon a plane boundary gives a reflected longitudinal wave but also excites a transverse reflected wave. For an ideal isotropic continuum, there is a boundary condition that avoids these complications—a perfect, smooth, and rigid boundary—but it does not work for crystal lattices. M. Born invented for that purpose a very ingenious type of condition which he characterizes as *cyclic condition*. He takes an infinite unbounded medium through which plane waves can propagate freely and selects inside this infinite medium a finite volume V , which is a rectangular parallelepiped.

$$0 \leq x \leq L_1, \quad 0 \leq y \leq L_2, \quad 0 \leq z \leq L_3, \\ V = L_1 L_2 L_3 \quad (39.3)$$

The condition for the waves is to take the same value at a point (x, y, z) and at points $(x + n_1 L_1)$, $(y + n_2 L_2)$, and $(z + n_3 L_3)$ where n_1 , n_2 , and n_3 are any arbitrary integers.

$$\psi(x + n_1 L_1, y + n_2 L_2, z + n_3 L_3) = \psi(x, y, z) \quad (39.4)$$

Since

$$\psi = e^{i\omega t - 2\pi i(ax + by + cz)} \quad n_1 n_2 n_3 \text{ integer} \quad (39.5)$$

this requires that

$$a = \frac{m_1}{L_1}, \quad b = \frac{m_2}{L_2}, \quad c = \frac{m_3}{L_3} \quad m_1 m_2 m_3 \text{ integer} \quad (39.6)$$

A characteristic frequency of the crystal will be excited whenever the wave vector terminates on a point with components given by Eq. (39.6). The number of such points per unit volume of the reciprocal lattice is $L_1 L_2 L_3$ (since the points are spaced at intervals $1/L_1$, $1/L_2$, and $1/L_3$ along the x , y , and z axes, respectively).

The number \mathcal{N} of atoms contained in the volume V is stupendous [Eq. (39.2)], and L_1 , L_2 , and L_3 are very large, compared with the dimensions d_1 , d_2 , and d_3 of the elementary lattice cell. This means that the density of the vibration points given by Eq. (39.6) is extremely high. Even if we take a very small volume element $d\tau$ in the abc space, we may obtain the average number of points inside $d\tau$ by just taking

$$dn = L_1 L_2 L_3 d\tau = V d\tau$$

Let us, for instance, consider a certain direction of propagation and a small cone of aperture $d\Omega$ around this average direction. The number of \mathbf{a} vectors that terminate inside this cone between the distances r and $r + dr$ is

$$dn = L_1 L_2 L_3 r^2 dr d\Omega = V r^2 dr d\Omega \quad (39.7)$$

$$r = |\mathbf{a}| = \frac{1}{\lambda} \quad (39.8)$$

Hence

$$dn = V d\Omega \frac{1}{\lambda^2} d\left(\frac{1}{\lambda}\right) = V d\Omega \frac{\nu^2}{W^2} d\left(\frac{\nu}{W}\right)$$

where W is the *phase velocity* for one special type of wave. According to the definitions given in Sec. 21, the *group velocity* U is defined by

$$W = \frac{\nu}{|\mathbf{a}|}, \quad U = \frac{d\nu}{d|\mathbf{a}|}, \quad \frac{1}{U} = \frac{d}{d\nu} \left(\frac{\nu}{W} \right) \quad (39.9)$$

and we finally obtain

$$dn = V d\Omega \frac{\nu^2 d\nu}{W^2 U} \quad (39.10)$$

In a crystal lattice the difficulty is that phase and group velocities

depend upon both ν and the orientation of the cone. In an ideal isotropic body this last dependence disappears, and one may immediately integrate about all orientations, thus replacing $d\Omega$ by 4π .

$$dn = 4\pi V \frac{\nu^2 d\nu}{W^2 l} \quad \text{isotropic in all directions} \quad (39.11)$$

There is, however, a general result that can be deduced directly from Eq. (39.7): For *each type of elastic wave*, the *total number of vibrations* (with cyclic conditions) is exactly *equal to the number \mathfrak{N} of atoms* in the volume V . This is easily seen, since the total number of vibrations of a certain type is obtained by integrating Eq. (39.7) over the whole first zone, the volume of which is V_b .

$$\left. \begin{aligned} \int r^2 dr d\Omega &= V_b \\ n &= \int dn = V V_b = \frac{V}{V_d} = \mathfrak{N} \end{aligned} \right\} \quad (39.12)$$

according to Eq. (39.1). This is a very important and general result of the zone theory.

How can this general scheme be simplified to be applied to the problem of an *ideal isotropic solid body*? The word "Ideal" is necessary to remind us of the somewhat artificial character of the assumptions: actual so-called *isotropic* solids are only mixtures of tiny crystals oriented at random. Our ideal isotropic

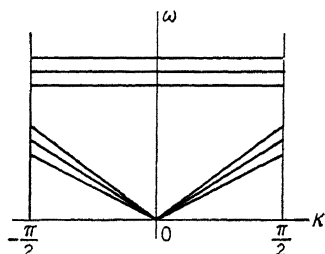


FIG. 39.1.

solid will represent a sort of average of crystal properties for the purpose of simplification. *Two assumptions* will be made:

a. We assume that the curves of Fig. 36.1 are replaced by straight lines, as shown in Fig. 39.1. This means that the upper optical branches are supposed to correspond to single frequencies (instead of the actual frequency bands of finite width) and that for the *acoustic branches* we assume a *constant phase velocity*

$$W = U = \text{const.} \quad (39.13)$$

b. We simplify the *shape of the first zone* and replace it by a *sphere*. But here we must not forget the general results emphasized at the beginning of this section, and we assume the *same*

spherical first zone for all waves (whether electromagnetic or longitudinal and transverse acoustical waves). Let the sphere have a radius R . Then the volume is

$$\frac{4}{3}\pi R^3 = V_b = \frac{1}{V_d} = \frac{\mathfrak{N}}{V} \quad (39.14)$$

where V_d is the volume of the cell in the direct lattice and \mathfrak{N} the number of atoms in a total volume V . Accordingly, we maintain the validity of Eq. (39.12) as necessary.

In order to comply with the requirements of the general theory we see that we have to introduce the same R for all types of waves, hence the same limit for the wave length.

$$\left. \begin{aligned} R &= |\mathbf{a}|_{\max} = \frac{1}{\lambda_{\min}} && \text{cutoff wave length} \\ \lambda_{\min} &= \left(\frac{4\pi V}{3\mathfrak{N}}\right)^{1/3} && \text{common to all waves} \end{aligned} \right\} \quad (39.15)$$

which means different cutoff frequencies for different waves. We may speak now of longitudinal waves (phase velocity W_l) and transverse waves (phase velocity W_t), and we obtain the maximum frequencies

$$\nu_{lc} = W_l \left(\frac{3\mathfrak{N}}{4\pi V}\right)^{1/3}, \quad \nu_{tc} = W_t \left(\frac{3\mathfrak{N}}{4\pi V}\right)^{1/3} \quad (39.16)$$

These assumptions are very close to the ones introduced by Debye in his famous theory of specific heats, but not quite the same. At the time of Debye's original paper, the theory of zones was not known, although some of the main results had already been obtained by Born. Hence, Debye did not realize the necessity of taking the same minimum wave length (and different cutoff frequencies) for the different waves. He found it easier to assume the same cutoff frequency and different cutoff wave lengths and stated the condition

$$\nu_D = \left(\frac{1}{W_l^3} + \frac{2}{W_t^3}\right)^{-1/3} \left(\frac{9\mathfrak{N}}{4\pi V}\right)^{1/3} \quad (39.17)$$

ν_D is Debye's single cutoff frequency. His reasoning was as follows: Starting from Eq. (39.11) and taking

$$\begin{aligned} W_l &= U_l = \text{const.} \\ W_t &= U_t = \text{const.} \end{aligned}$$

he first computed the total number of vibrations in the frequency interval $(\nu, \nu + d\nu)$ for all waves (one longitudinal and two transverse) and wrote

$$dn_l + 2dn_t = 4\pi V \nu^2 d\nu \left(\frac{1}{W_l^3} + \frac{2}{W_t^3} \right) \quad (39.18)$$

Then he integrated from $\nu = 0$ to ν_D and assumed that the total number of vibrations was $3\mathfrak{N}$. Hence

$$3\mathfrak{N} = \frac{4}{3} \pi V \nu^3 \left(\frac{1}{W_l^3} + \frac{2}{W_t^3} \right) \quad (39.19)$$

which is the same as Eq. (39.17). The weak point in Debye's argument is that this oversimplification does not satisfy the

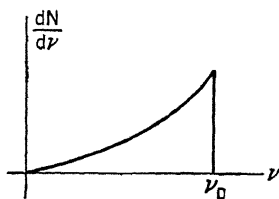


FIG. 39.2.

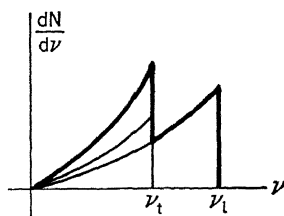


FIG. 39.3.

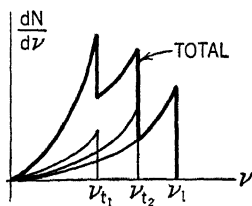


FIG. 39.4.

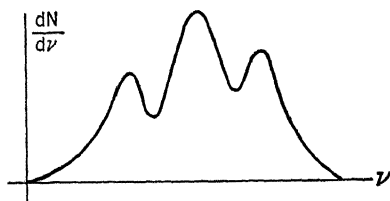


FIG. 39.5.

necessary condition (39.12). We must always have \mathfrak{N} longitudinal and $2\mathfrak{N}$ transverse vibrations. With Debye's assumption the distribution between longitudinal and transverse modes is modified; only the total number is maintained by $3\mathfrak{N}$.

Figure 39.2 shows Debye's distribution function, while Fig. 39.3 shows ours. In Fig. 39.3 the two curves are to be added (the sum is given by the heavy curve) to obtain the actual distribution. Figure 39.4 shows the curves that would be obtained for an anisotropic medium. The sharp points on our curves are

due to the assumption of a spherical first zone. A better approximation yields a curve of the type shown in Fig. 39.5.

40. The Energy of a Solid; the Characteristic Temperatures

We regard the solid as made up of $3\mathfrak{N}$ harmonic oscillators. The energy of a harmonic oscillator vibrating with frequency ν is, according to Planck (quantum theory),

$$u_\nu = \frac{h\nu}{e^{h\nu/kT} - 1} + \frac{1}{2} h\nu \quad (40.1)$$

where k is Boltzmann's constant, h is Planck's constant, and T is the absolute temperature. The constant $\frac{1}{2}h\nu$ did not appear in the original Planck treatment, but quantum mechanical considerations indicate that this, and not zero, is the correct energy at $T = 0$. Each longitudinal or transverse vibration of frequency ν receives an amount of energy u_ν , and the total energy of the ideal solid at temperature T is

$$U_T = \sum_{m=l,t_1,t_2} \int u_\nu dn_m \quad (40.2)$$

The upper limit of integration is the cutoff frequency for the corresponding type of wave.

These frequencies can be used to define characteristic temperatures. A first definition Θ_D is the one of Debye and is based upon his single frequency limit ν_D .

$$k\Theta_D = h\nu_D \quad \text{or} \quad \Theta_D = \frac{h\nu_D}{k} \quad (40.3)$$

Our theory will evidently yield two characteristic temperatures, since the cutoff frequencies are different for longitudinal and transverse waves [Eq. (39.16)].

$$\left. \begin{aligned} \Theta_l &= \frac{h\nu_l}{k} = \frac{hW_l}{k} \left(\frac{3\mathfrak{N}}{4\pi V} \right)^{1/3} \\ \Theta_t &= \frac{h\nu_t}{k} = \frac{hW_t}{k} \left(\frac{3\mathfrak{N}}{4\pi V} \right)^{1/3} \end{aligned} \right\} \quad (40.4)$$

If we denote the type of wave by the subscript m , the total energy at temperature T is, according to Eqs. (39.11), (40.1), and (40.2),

$$U_T = \sum_{m=l_1, l_2} \frac{4\pi V}{W_m^3} \int_0^{\nu_m} \nu^2 d\nu \left[\frac{h\nu}{e^{h\nu/kT} - 1} + \frac{1}{2} h\nu \right] \quad (40.5)$$

We make the following change of variables:

$$\xi = \frac{h\nu}{kT}, \quad \xi_m = \frac{h\nu_m}{kT} = \frac{\Theta_m}{T}$$

and Eq. (40.5) becomes

$$U_T = \sum_{l, l_1, l_2} \frac{4\pi V}{W_m^3} \frac{(kT)^4}{h^3} \int_0^{\xi_m} \xi^3 d\xi \left[\frac{1}{e^\xi - 1} + \frac{1}{2} \right] \quad (40.6)$$

This integration must be performed by approximate methods for most cases. We define the Debye function

$$D(x) = \frac{3}{x^3} \int_0^x \left[\frac{1}{e^\xi - 1} + \frac{1}{2} \right] \xi^3 d\xi \quad (40.7)$$

Now the coefficient of the integral in Eq. (40.6) is given by

$$\frac{4\pi V}{W_m^3} \frac{(kT)^4}{h^3} \quad (40.8)$$

and, if we multiply by $\xi_m^3/3$, we can replace the integral by the Debye function. Then

$$\begin{aligned} \frac{4\pi V}{W_m^3} \frac{(kT)^4}{h^3} \frac{\xi_m^3}{3} &= \frac{4\pi V}{3W_m^3} \frac{(kT)^4}{h^3} \frac{h^3 \nu_m^3}{k^3 T^3} \\ &= \frac{4\pi V}{3} \frac{\nu_m^3}{W_m^3} kT = \frac{4\pi V}{3} kT \left(\frac{3\mathcal{N}}{4\pi V} \right) = \mathcal{N}kT = RT \end{aligned}$$

Thus we may write Eq. (40.6) in the form

$$U_T = \sum_m U_{T_m} = RT \sum_{l, l_1, l_2} D\left(\frac{\Theta_m}{T}\right)$$

and, if the two transverse waves are alike, this becomes

$$U_T = RT \left[D\left(\frac{\Theta_l}{T}\right) + 2D\left(\frac{\Theta_t}{T}\right) \right] \quad (40.9)$$

The Debye treatment gives

$$U_T = 3RTD\left(\frac{\Theta_D}{T}\right) \quad (40.10)$$

In general, Eq. (40.9) is more reliable than Eq. (40.10). However, the two characteristic temperatures Θ_l and Θ_t are often both fairly close to Θ_D and for very high or very low temperatures the two theories agree.

For very high temperatures, classical thermodynamics holds with equipartition of energy, as can be seen from Eq. (40.1).

$$u_r = kT \left(\frac{\xi}{e^{\xi} - 1} + \frac{1}{2} \xi \right) = kT \left(\frac{\xi}{\xi + \frac{1}{2}\xi^2 + \dots} + \frac{1}{2} \xi \right) \\ \approx kT \left(1 - \frac{\xi}{2} + \frac{\xi}{2} \right) = kT \quad (40.11)$$

Each of the oscillators has an energy kT . Both theories have been formulated in such a way as to yield the correct number $3\mathcal{N}$ for the total number of degrees of freedom; hence we obtain a total energy $3\mathcal{N}kT$ or $3RT$ in both cases. This may also be seen from Eq. (40.7), since for very high temperatures Debye's function approaches one and both theories yield the result

$$U_T = 3RT \quad (40.12)$$

which is just the law of Dulong and Petit. For temperatures so small that

$$T \ll \Theta_e, \Theta_D, \Theta_t$$

holds, Θ_m/T is very large, and the upper limit in the integral of Eq. (40.5) may be replaced by ∞ . Physically this means that only the low frequencies are excited and the number and value of the higher frequencies is unimportant. The Debye function becomes

$$\lim_{x \rightarrow \infty} D(x) = \lim_{x \rightarrow \infty} \frac{3}{x^3} \int_0^x \left(\frac{1}{e^{\xi} - 1} + \frac{1}{2} \right) \xi^3 d\xi \\ = \lim_{x \rightarrow \infty} \frac{3}{x^3} \left(\frac{x^4}{8} + \int_0^{\infty} \frac{1}{e^{\xi} - 1} \xi^3 d\xi \right) \quad (40.13)$$

The last integral is a constant that is found equal to $\pi^4/15$.

$$U_T = \sum_m \left[RT \frac{3}{8} \frac{\Theta_m}{T} + 3RT \frac{\pi^4}{15} \left(\frac{T}{\Theta_m} \right)^3 \right] \\ = \frac{3}{8} R (\Theta_l + 2\Theta_t) + RT \frac{\pi^4}{5} \left[\left(\frac{T}{\Theta_l} \right)^3 + 2 \left(\frac{T}{\Theta_t} \right)^3 \right] \quad (40.14)$$

The first term is a constant and will not enter into the applications to specific heats, etc., since these involve derivatives of U_T . Debye's solution for U_T is

$$U_T = 3 \frac{\pi^4}{5} RT \left(\frac{T}{\Theta_D} \right)^3 + C \quad (40.15)$$

where C is a constant, and so the two theories agree for low temperatures, on account of the relation

$$\frac{1}{\Theta_l^3} + \frac{2}{\Theta_l^3} = \frac{3}{\Theta_D^3} \quad (40.16)$$

which is easily obtained from Eqs. (40.3), (40.4), and (39.17).

41. Thermal Expansion and Entropy of a Solid Body

We close our treatment of solids with a discussion of the entropy. To do this, we must take account of the radiation pressure due to elastic waves. The radiation pressure is due to deviations from Hooke's law and has been computed by Rayleigh and L. Brillouin:

$$p_m = \frac{U_m}{V} \left(\frac{1}{3} - \frac{V}{W_m} \frac{\partial W_m}{\partial V} \right) \quad (41.1)$$

where U_m denotes the total energy of the waves of type m . For electromagnetic radiation in vacuum,

$$p_m = \frac{1}{3} \frac{U_m}{V}$$

since in this case there is no $\frac{\partial W_m}{\partial V}$ term. Equation (41.1) may also be written in the form

$$p_m = - \frac{U_m}{\Theta_m} \frac{\partial \Theta_m}{\partial V} \quad (41.2)$$

since from Eq. (40.4)

$$\begin{aligned} \frac{\partial \log \Theta_m}{\partial V} &= \frac{1}{\Theta_m} \frac{\partial \Theta_m}{\partial V} = \frac{\partial}{\partial V} \left(\log \frac{h}{k} W_m - \frac{1}{3} \log \frac{4\pi V}{3\mathfrak{N}} \right) \\ &= \frac{1}{W_m} \frac{\partial W_m}{\partial V} - \frac{1}{3V} \end{aligned}$$

or

$$- \frac{V}{\Theta_m} \frac{\partial \Theta_m}{\partial V} = \frac{1}{3} - \frac{V}{W_m} \frac{\partial W_m}{\partial V}$$

The external pressure must balance the internal pressure. The latter is given by the sum of the pressure due to the interactions of the molecules of the crystal and the radiation pressure.

$$p = f(V) + \sum_m p_m \quad (41.3)$$

If the external pressure is zero, then

$$f(V) = - \sum_m p_m$$

and we may think of the radiation pressure as doing work against the cohesive forces of the molecules during an expansion due to rise in temperature.

The total energy of the solid is

$$U = F(V) + U_1 + U_{t_1} + U_{t_2} \quad (41.4)$$

where

$$F(V) = - \int f(V) dV$$

During a small expansion the work done is

$$\begin{aligned} dW &= p dV = f dV - \sum_m \frac{U_m}{\Theta_m} \frac{\partial \Theta_m}{\partial V} dV \\ &= f dV - RT \sum_m \left[D \left(\frac{\Theta_m}{T} \right) \frac{1}{\Theta_m} \frac{\partial \Theta_m}{\partial V} dV \right] \end{aligned} \quad (41.5)$$

where U_m is assumed to be given by Eq. (40.9). The change in energy is given by

$$\begin{aligned} dU &= \frac{\partial U}{\partial V} dV + \frac{\partial U}{\partial T} dT = -f dV + \sum_m \frac{\partial U_m}{\partial V} dV + \sum_m \frac{\partial U_m}{\partial T} dT \\ &= -f dV + R \sum_m \left[D' \left(\frac{\Theta_m}{T} \right) \frac{\partial \Theta_m}{\partial V} dV + D dT - \frac{\Theta_m}{T} D' \left(\frac{\Theta_m}{T} \right) dT \right] \end{aligned} \quad (41.6)$$

Now the heat change will be

$$dQ = dU + dW$$

and the change in entropy

$$dS = \frac{dQ}{T} = \frac{dU + dW}{T} \quad (41.7)$$

Substituting Eqs. (41.5) and (41.6) into Eq. (41.7) gives

$$\begin{aligned} dS &= \frac{R}{T} \sum_m \left[D dT - D \frac{T}{\Theta_m} \frac{\partial \Theta_m}{\partial V} dV + D' \left(\frac{\partial \Theta_m}{\partial V} dV - \frac{\Theta_m}{T} dT \right) \right] \\ &= R \sum_m \left(\frac{D'}{T} - \frac{D}{\Theta_m} \right) \left(d\Theta_m - \frac{\Theta_m}{T} dT \right) \end{aligned}$$

on replacing $\frac{\partial \Theta_m}{\partial V} dV$ by $d\Theta_m$. If we let $X_m = \Theta_m/T$

$$dS = R \sum_m \left(D' - \frac{D}{X_m} \right) dX_m$$

and

$$\frac{dS}{dX_m} = R \left(D' - \frac{D}{X_m} \right) \quad (41.8)$$

We can write this as an explicit function of X_m by using Eq. (40.7)

$$D(X_m) = \frac{3}{X_m^3} \int_0^{X_m} \frac{\xi^3 d\xi}{e^\xi - 1} + \frac{3}{8} X_m$$

in Eq. (41.8), which then becomes

$$\begin{aligned} \frac{dS}{dX_m} &= R \left(-\frac{9}{X_m^4} \int_0^{X_m} \frac{\xi^3 d\xi}{e^\xi - 1} + \frac{3}{e^{X_m} - 1} - \frac{3}{X_m^4} \int_0^{X_m} \frac{\xi^3 d\xi}{e^\xi - 1} \right) \\ &= \frac{3R}{e^{X_m} - 1} - \frac{12R}{X_m^4} \int_0^{X_m} \frac{\xi^3 d\xi}{e^\xi - 1} \end{aligned} \quad (41.9)$$

It is interesting to note that Eq. (41.9) agrees with the quantum mechanical result for the entropy of our system of harmonic oscillators. The quantum theory offers a possibility for a direct statistical computation of entropy. A distribution of n quanta $h\nu$ (total energy $n h\nu$) among g resonators is found to have an entropy

$$S = kg \left[\left(1 + \frac{n}{g} \right) \log \left(1 + \frac{n}{g} \right) - \frac{n}{g} \log \frac{n}{g} \right] \quad (41.10)$$

The most probable distribution corresponds to Planck's formula for the average energy per resonator.

$$u_\nu = \frac{n h\nu}{g} = \frac{h\nu}{e^{h\nu/kT} - 1} \quad \xi = \frac{h\nu}{kT} \quad (41.11)$$

Hence

$$\frac{n}{g} = \frac{1}{e^\xi - 1} \quad 1 + \frac{n}{g} = \frac{e^\xi}{e^\xi - 1}$$

On the other hand, we have obtained an expression (39.11) for the number of resonators of frequency $\nu(d\nu)$. Using this value for g ,

$$g = V \frac{4\pi\nu^2 d\nu}{15\pi^3} \quad \text{wave type } m \quad (41.12)$$

and integrating on ν from 0 to the limit ν_m , we obtain the total entropy of the solid.

$$S = \sum_{m=1,2,3} S_m \quad (41.13)$$

$$S_m = 3\Re k \left(\frac{T}{\Theta_m} \right)^3 \int_0^{\Theta_m/T} \left(\frac{e^\xi}{e^\xi - 1} \log \frac{e^\xi}{e^\xi - 1} - \frac{1}{e^\xi - 1} \log \frac{1}{e^\xi - 1} \right) \xi^2 d\xi$$

The method followed for entropy is exactly the same as the one used for energy, in Eqs. (40.2) to (40.9). Some regrouping of terms and elementary transformations finally yields

$$S_m = \frac{3R}{X_m^3} \int_0^{X_m} \left[\frac{\xi e^\xi}{e^\xi - 1} - \log(e^\xi - 1) \right] \xi^2 d\xi \quad (41.14)$$

We may integrate by parts to eliminate the logarithmic term.

$$-\int_0^{X_m} \log(e^\xi - 1) \xi^2 d\xi = -\frac{X_m^3}{3} \log(e^{X_m} - 1) + \frac{1}{3} \int_0^{X_m} \frac{e^\xi}{e^\xi - 1} \xi^3 d\xi$$

so that

$$S = \sum_m \left[-R \log(e^{X_m} - 1) + \frac{4R}{X_m^3} \int_0^{X_m} \frac{e^\xi \xi^3 d\xi}{e^\xi - 1} \right] \quad (41.15)$$

Differentiation with respect to X_m yields

$$\begin{aligned} \frac{dS}{dX_m} &= -\frac{Re^{X_m}}{e^{X_m} - 1} - \frac{12R}{X_m^4} \int_0^{X_m} \frac{e^\xi \xi^3 d\xi}{e^\xi - 1} + \frac{4Re^{X_m}}{e^{X_m} - 1} \\ &= \frac{3Re^{X_m}}{e^{X_m} - 1} - \frac{12R}{X_m^4} \int_0^{X_m} \frac{e^\xi \xi^3 d\xi}{e^\xi - 1} \end{aligned} \quad (41.16)$$

This is easily seen to check with Eq. (41.9) by the following transformation:

$$\frac{dS}{dX_m} = 3R \left(\frac{1}{e^{X_m} - 1} + 1 \right) - \frac{12R}{X_m^4} \int_0^{X_m} \left(\frac{1}{e^\xi - 1} + 1 \right) \xi^3 d\xi$$

The (+1) term in the second bracket yields, after integration,

$$-\frac{12R}{X_m^4} \frac{\xi^4 X_m}{4} \Big|_0 = -3R$$

which cancels the (+1) term in the first bracket and yields Eq. (41.9).

Thus the whole theory is proved to be entirely consistent. Debye's theory with its single characteristic temperature does not work, since it is not consistent with Eq. (41.2) for internal pressure.

Further details on the theory of solids and the direct computation of Eq. (41.1) for the radiation pressure can be found in the author's book, "Les Tenseurs en mécanique et en élasticité."¹ This short summary of the problem was intended to emphasize the connection of this problem with the theory of zones, as explained in Secs. 39 and 40, and to show the necessity of a correction to the original Debye theory involving the use of two characteristic temperatures, as pointed out by Born in his discussion of the theory of solids.

Selected References on Zones

Books

- BRILLOUIN, L.: "Quantenstatistik," Springer, Berlin, 1931.
 BRILLOUIN, L.: "Structure des corps solides," Hermann, Paris, 1937.
 MOTT, N. F., and H. JONES: "Theory of Metals and Alloys," Oxford, New York, 1936.
 SEITZ, F.: "The Modern Theory of Solids," McGraw-Hill, New York, 1940.

Original papers

- BLOCH, F.: *Z. Physik*, **52**, 555 (1928).
 MORSE, P. M.: *Phys. Rev.*, **35**, 1310 (1930).
 PEIERLS, R.: *Ann. Physik*, **4**, 121 (1930).
 BRILLOUIN, L.: *J. phys.*, **1**, 377 (1930); **3**, 565 (1932); **4**, 1, 333 (1933); **7**, 401 (1936).
 JONES, H.: *Proc. Roy. Soc. (London)*, **A144**, 225 (1934); **A147**, 396 (1934).
 JONES, H., N. F. MOTT, and H. W. B. SKINNER: *Phys. Rev.*, **45**, 379 (1934).
 SLATER, J. C.: *Phys. Rev.*, **45**, 794 (1934).
 HUND, F.: *Z. tech. Physik*, **16**, 331, 494 (1935). *Z. Physik*, **99**, 119 (1936).
 PEIERLS, R.: *Ann. Inst. Henri Poincaré*, **5**, 177 (1935).
 SEITZ, F.: *Ann. Math.*, **37**, 17 (1936).
 BOUCKAERT, L. P., R. SMOLUCHOWSKI, and E. WIGNER: *Phys. Rev.*, **50**, 58 (1936).
 HERRING, C.: *Phys. Rev.*, **52**, 361, 365 (1937).
 CHODOROW, M. I., and M. F. MANNING: *Phys. Rev.*, **52**, 731 (1937).

¹ Masson, Paris, 1938.

CHAPTER VIII

MATHIEU'S EQUATION AND RELATED PROBLEMS

42. Mathieu's Equation

The problem of the propagation of waves in a periodic continuous medium was discussed in Sec. 34, and there some general results were obtained on the type of the solution and its properties. Few examples of such problems have been completely discussed. Some of the best known relate to the one-dimensional problem. The general equation given in Eq. (34.1) reduces to

$$\left. \begin{aligned} \frac{\partial^2 \psi}{\partial x^2} + \omega^2 F(x) \psi &= 0 \\ \psi &= e^{i\omega t} u(x) \end{aligned} \right\} \quad (42.1)$$

Mathieu's equation is obtained when the periodic function F (of period d in x) contains only one cosine term and the expansion (34.3) can be written

$$F(x) = C_0 + C_1 e^{2\pi i(x/d)} + C_1 e^{-2\pi i(x/d)} = C_0 + 2C_1 \cos \frac{2\pi x}{d} \quad (42.2)$$

So far we have discussed equations of this type in detail only for very small C_1 (corresponding to a very small perturbation) when we assumed from Eqs. (35.1) and (35.3)

$$F = C_0 + \epsilon f \quad \epsilon \text{ very small}$$

Now we wish to discuss the solutions for any arbitrary value of both C_0 and C_1 . We may use the following notation to reduce the equation to standard type: We introduce a new variable

$$\xi = \frac{\pi x}{d} \quad (42.3)$$

which has period π instead of d , and we obtain

$$\left. \begin{aligned} \frac{\partial^2 u}{\partial \xi^2} + (\eta + \gamma \cos 2\xi) u &= 0 \\ \eta = \omega^2 \frac{d^2}{\pi^2} C_0 = 4\nu^2 d^2 C_0 &\quad \gamma = 8\nu^2 d^2 C_1 \end{aligned} \right\} \quad (42.4)$$

This equation was first obtained by Mathieu in connection with the problem of vibrations of an elliptic membrane, and this is the reason for the choice of π as the period. The same equation is found for the oscillations of an elliptic lake, for the tides in an elliptic sea, and, in general, in all problems concerning waves or vibrations with an elliptic boundary, such as propagation of acoustic or electromagnetic waves along a pipe of elliptic cross section. We shall indicate in a later section some other interesting problems of physics and engineering where Mathieu's equation plays a major role.¹

Floquet discovered that the general solution of the equation could be written

$$u = D_1 A(\xi) e^{\mu \xi} + D_2 A(-\xi) e^{-\mu \xi} \quad (42.5)$$

with an amplitude $A(\xi)$ that is a periodic function of ξ with period π . This solution is thus a superposition of two waves propagated (or attenuated) in opposite directions. This is clearly seen if we retain the $e^{i\omega t}$ factor and write the original ψ function of Eq. (42.1). D_1 and D_2 are arbitrary constants. If we keep only *one* of these waves, we obtain

$$u = A(\xi) e^{\mu \xi} \quad A(\xi) \text{ has period } \pi. \quad (42.6)$$

This corresponds exactly to our general solution (34.4).

In our previous discussion we assumed that we had to deal with *actual waves*, and we therefore took

$$\mu = i\beta = i \frac{k}{\pi}, \quad \mu \xi = i\beta \xi = i \frac{kx}{d} = i2\pi ax$$

and determined afterwards the corresponding ω value and the frequency. Here the discussion is conducted in the opposite way. The frequency has been chosen as primary data, and the problem is to obtain μ , which may be

$$\left. \begin{array}{ll} \mu = i\beta & \text{pure imaginary, unattenuated sine waves} \\ \mu = \alpha + i\beta & \text{complex or real, attenuated motion} \end{array} \right\} \quad (42.7)$$

Both η and γ are proportional to ω^2 [from Eq. (42.4)], and hence increasing frequency means increasing η and γ . Their ratio is constant. The diagram in Fig. 42.1 summarizes the results.

¹ Different authors writing on Mathieu functions use widely different notations: this text, η , γ ; Mathieu, R , $\pm 2h^2$; Strutt, λ , $2h^2$; Strutt and van der Pol, $4\omega^2$, $4\alpha^2$; Whittaker and Humbert, a , $16q$.

The coordinates are η and γ . The plain white regions correspond to (η, γ) values for which μ is complex or real and attenuation occurs. The shaded regions are those for which μ is pure imaginary and yields unattenuated waves. In the language of the engineers, blank regions mean absorption or stopping bands and shaded regions mean passing bands.

The two lines $\eta = \gamma$ and $\eta = -\gamma$ are drawn as guides. All the passing bands become straight lines parallel to $\eta = -\gamma$ at

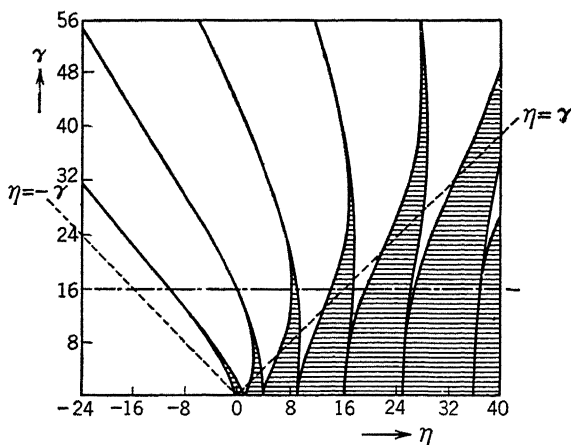


FIG. 42.1.

infinity. These lines intersect a line parallel to the η axis at the point given by

$$\eta = -\gamma + (2n + 1) \sqrt{2\gamma} \quad (42.8)$$

if the line is drawn sufficiently far above the η axis. Evidently, then, they also become parallel to one another at infinity.

There is no propagation for any value of $\eta < -\gamma$. We can make the following statements about the character of the waves:

- | | | |
|------------------------------|---|----------|
| 1. $\eta < -\gamma$, | μ complex or real, absorption | } (42.9) |
| 2. $-\gamma < \eta < \gamma$ | $\left\{ \begin{array}{l} \mu \text{ real or complex, broad} \\ \text{absorption bands} \\ \mu \text{ pure imaginary, narrow} \\ \text{transmission bands} \end{array} \right.$ | |
| 3. $\eta > \gamma$ | $\left\{ \begin{array}{l} \mu \text{ real or complex, narrow} \\ \text{absorption bands} \\ \mu \text{ pure imaginary, broad trans-} \\ \text{mission bands} \end{array} \right.$ | |

The boundary between blank and shaded regions is the curve

$$\text{R.P. } \mu = 0 \quad (\text{R.P. means "real part of."}) \quad (42.10)$$

The η axis corresponds to $\gamma = 0$ and hence to a continuous uniform medium with no periodic structure whatever. The entire positive part of the η axis lies inside the shaded region, since our definitions in Eq. (42.4) mean, in case of propagation,

$$C_0 = \frac{1}{V_0^2}, \quad \eta = \frac{4\pi^2 d^2}{V_0^2} > 0 \quad (42.11)$$

A small perturbation is obtained when $\gamma \ll \eta$, i.e., in the immediate neighborhood of the η axis. The boundary curves (42.10) leave the η axis at the points

$$\eta = n^2 \quad n = \text{integer} \quad (42.12)$$

and have contacts of order $(n - 1)$ at these points:

$n = 1$, two curves crossing each other

$n = 2$, two curves with a common vertical tangent

$n = 3$, a common vertical tangent and the same curvature, etc.

These curves were very carefully computed by Mathieu himself and checked by later computations.

43. Mathieu Functions: General Discussion

The general solution, according to Floquet, was written in Eq. (42.5), and in this discussion emphasis can be laid either upon the imaginary exponentials [free waves as in Eq. (42.6)] or upon the real combinations that correspond to standing waves. This latter type of solution is the one mostly used by mathematicians who have computed the numerical solutions of Mathieu's equation. They start from the solution obtained along the η axis ($\gamma = 0$, no periodic perturbation), which they write

$$\cos m\xi, \quad \sin m\xi, \quad \mu_0 = im = i\sqrt{\eta} \quad (43.1)$$

In this case, $A(\xi)$ is a constant, and the correspondence with Eq. (42.5) is contained in the well-known relations

$$\begin{aligned} \cos m\xi &= \frac{1}{2}(e^{im\xi} + e^{-im\xi}) & D_1 &= D_2 = \frac{1}{2} \\ \sin m\xi &= \frac{1}{2i}(e^{im\xi} - e^{-im\xi}) & D_1 &= -D_2 = \frac{1}{2i} \end{aligned}$$

For $\gamma \neq 0$, Mathieu functions have been defined:

$$Ce_m(\gamma, \xi) \quad \text{and} \quad Se_m(\gamma, \xi) \quad (43.2)$$

They are known as expansions in powers of γ , starting with either $\cos m\xi$ or $\sin m\xi$ as the independent term. Each of these functions corresponds to the same value of m , hence to the same μ .

$$\mu = im = f(\eta, \gamma), \quad \gamma \rightarrow 0, \quad m \rightarrow \sqrt{\eta} \quad (43.3)$$

Lines of given μ (or m) run through the shaded areas of Fig. 42.1, as schematically shown in Fig. 43.1. On each line, two functions Ce_m and Se_m may be computed, except on the boundaries of the shaded area. On these boundaries only one of the two functions

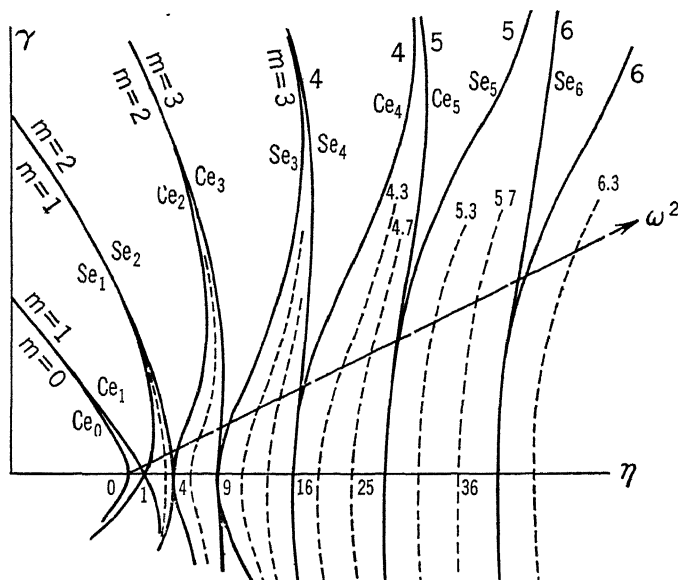


FIG. 43.1.

is obtained, as shown in Fig. 43.1. The boundaries correspond to integral values for m according to Eqs. (42.12) and (43.1) and start from the point $\eta = m^2 = 1, 4, 9, 16, \dots$ on the η axis. Two m curves start from each of these points: one curve yields the function Se_m , and the other curve corresponds to Ce_m . On each curve there is an additional aperiodic solution.

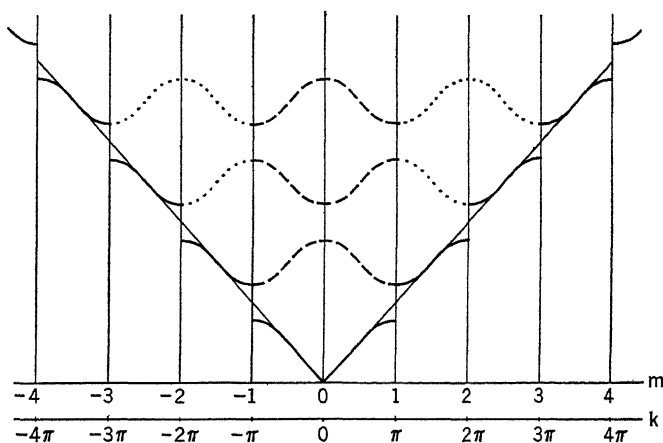
From these diagrams we can see how to obtain information about waves propagating through a periodic medium. The medium is defined by two constant coefficients C_0 and C_1 , and Eq. (42.4) yields η and γ each proportional to ν^2 , with a fixed ratio

$$\frac{\eta}{\gamma} = \frac{C_0}{2C_1} \quad (43.4)$$

This means a straight line running from the origin as shown in Fig. 43.1. The line cuts the successive curves corresponding to different m values. The shaded area of Fig. 42.1 gives passing bands, and the blank areas give stopping bands. The correspondence with our former notation results from Eqs. (42.6) and (43.3).

$$i\mu = -k \frac{1}{\pi} = -m \quad (43.5)$$

Hence $+m\pi$ plays the same role as our former k . The type of the resulting (ν, m) curve is shown in Fig. 43.2. Here, again,



FIRST ZONE

FIG. 43.2.

instead of choosing k (or m) values that insure continuity with the unperturbed uniform medium ($\gamma = 0$), we may select each time the smallest $|k|$ and reduce all curves inside the first zone $-\pi \leq k \leq \pi$. The important point is that *increased perturbation* (increased C_1) *does not modify the position* of the discontinuities but only *increases the magnitude* of the discontinuities.

The general theory of Mathieu's equation is very thoroughly discussed in Whittaker and Watson's "Modern Analysis,"¹ which contains some important results concerning the infinite determinants that we obtained in Sec. 34 in the preceding chap-

¹ Chap. IX, Sec. 19.41.

ter. Whittaker considers Hill's equation, which is a generalization of Mathieu's and which he writes as

$$\frac{d^2 u}{dx^2} + J(x)u = 0 \quad (43.6)$$

where J is a periodic function (period π) of x . This is our general equation of Chap. VII, reduced to one dimension. The correspondence of notations is as follows:

$$\left. \begin{array}{lllll} \text{Brillouin:} & d & b = \frac{1}{d} & 2\pi a & \omega^2 C_n & \omega^2 F \\ \text{Whittaker and} & & & & & \\ \text{Watson:} & \pi & \frac{1}{\pi} & i\mu & \theta_n & J \end{array} \right\} \quad (43.7)$$

and Eq. (43.6) reduces to our former Eq. (42.1).

The discussion of Secs. 34 and 35 centered on the determinant given in Eqs. (34.14) and (34.15).

$$\Delta = |\Delta_{m,p}| = \left| \frac{C_{m-p}}{(a - mb)^2} - \frac{1}{\nu^2} \delta_{mp} \right| \quad (43.8)$$

which according to Eq. (43.7) corresponds to the determinant with the elements

$$\Delta_{m,p} = \frac{1}{\nu^2} \left[\frac{\omega^2 C_{m-p}}{(i\mu - 2m)^2} - \delta_{mp} \right] \quad (43.9)$$

Whittaker and Watson assume that the θ_n series is absolutely convergent and compute another determinant

$$\left. \begin{array}{l} \Delta_1(i\mu) = |B_{mp}| \\ B_{mp} = - \frac{\theta_{r-p}}{(2m - i\mu)^2 - \theta_0} \quad m \neq p \\ B_{mm} = 1 \end{array} \right\} \quad (43.10)$$

These elements can be represented by the general formula

$$B_{mp} = \frac{-\theta_{m-p} + \delta_{mp}(2m - i\mu)^2}{-\theta_0 + (2m - i\mu)^2} = \nu^2 \Delta_{mp} \frac{(2m - i\mu)^2}{\omega^2 C_0 - (2m - i\mu)^2} \quad (43.11)$$

which gives the correspondence with our former notation. The determinant Δ_1 of Whittaker and Watson is shown to be equal to

$$\Delta_1(i\mu) = 1 + K \left[\cot \frac{\pi}{2} (i\mu + \sqrt{\theta_0}) - \cot \frac{\pi}{2} (i\mu - \sqrt{\theta_0}) \right] \quad (43.12)$$

where the constant K can be obtained by computing $\Delta_1(0)$.

$$\Delta_1(i\mu) = 1 - K \frac{\sin \pi \sqrt{\theta_0}}{\sin \frac{\pi}{2} (i\mu + \sqrt{\theta_0}) \sin \frac{\pi}{2} (i\mu - \sqrt{\theta_0})}$$

$$\Delta_1(0) = 1 + 2K \cot \frac{\pi}{2} \sqrt{\theta_0},$$

$$K = \frac{\Delta_1(0) - 1}{2 \cot \frac{\pi}{2} \sqrt{\theta_0}} = \frac{\Delta_1(0) - 1}{2} \tan \frac{\pi}{2} \sqrt{\theta_0}$$

Hence

$$\Delta_1(i\mu) = 1 - [\Delta_1(0) - 1] \frac{\sin^2 \frac{\pi}{2} \sqrt{\theta_0}}{\sin \frac{\pi}{2} (i\mu + \sqrt{\theta_0}) \sin \frac{\pi}{2} (i\mu - \sqrt{\theta_0})} \quad (43.13)$$

This enables one to write the fundamental equation stating that the determinant is zero.

$$\sin \frac{\pi}{2} (i\mu + \sqrt{\theta_0}) \sin \frac{\pi}{2} (i\mu - \sqrt{\theta_0}) = [\Delta_1(0) - 1] \sin^2 \frac{\pi}{2} \sqrt{\theta_0} \quad (43.14)$$

or

$$-\cos^2 \left(\frac{\pi}{2} i\mu \right) + \cos^2 \left(\frac{\pi}{2} \sqrt{\theta_0} \right) = [\Delta_1(0) - 1] \sin^2 \frac{\pi}{2} \sqrt{\theta_0}$$

Hence

$$\sin^2 \left(\frac{\pi}{2} i\mu \right) = \Delta_1(0) \sin^2 \frac{\pi}{2} \sqrt{\theta_0} \quad (43.15)$$

which is Whittaker's result, except that he introduces another determinant $\Delta(i\mu)$ that is equal to $\Delta_1(0)$ for $\mu = 0$.

These relations are very interesting since they apply to Hill's general equation and not only to the Mathieu problem. An attempt to find an extension of these results for the three-dimensional problem of Sec. 34 would, if successful, be very important.

From our former notation, according to Eq. (43.7), Whittaker's equation (43.15) becomes

$$\sin^2 (\pi^2 a) = \Delta_1(0) \sin^2 \frac{\pi}{2} \omega \sqrt{C_0} = \Delta_1(0) \sin^2 \left(\frac{\pi^2 \nu}{V_0} \right) \quad (43.16)$$

and yields the general relation between wave number a and frequency ν for a one-dimensional lattice of period $d = \pi$ in x . The fact that ν appears as a periodic function of period $1/d = 1/\pi$ in a is obvious, as are the discontinuities in the $\nu(a)$ relation when $\Delta_1(0) > 1$. The similarity of this equation with those obtained in Chap. III is also very striking.

44. Hill's Equation with a Rectangular Curve

As already stated, Hill's equation is obtained when the cosine term in Mathieu's equation is replaced by an arbitrary periodic function of x . Let us write it as

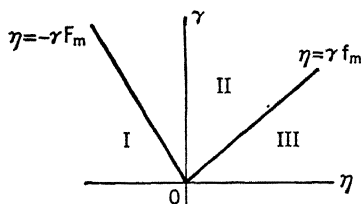


FIG. 44.1.

$$\frac{\partial^2 u}{\partial x^2} + [\eta + \gamma f(x)]u = 0 \quad (44.1)$$

Strutt has shown that some general results about this equation can be obtained. Assuming the integral of f over a period to be zero so as to make

$$\bar{f} = 0$$

and calling f_m and $-f_m$ the maximum and minimum values of f , respectively, it is possible to show that the $\eta\gamma$ plane is again

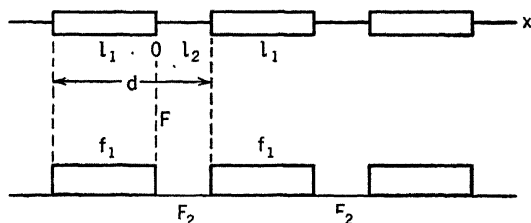


FIG. 44.2.

divided into three regions (Fig. 44.1) by the lines

$$\eta = -\gamma f_m, \quad \eta = \gamma f_m \quad (44.2)$$

In the first region μ is complex. In region II, μ is generally complex with some narrow bands where it becomes pure imaginary. In region III, μ is generally imaginary with some narrow bands where it is complex.

This may be shown in an example where the computations can be carried out explicitly. Let us assume a line built as shown in Fig. 44.2, with alternate portions of lengths l_1 and l_2 with values f_1 and f_2 . This is obviously a special case of Eq. (44.1) with the function f represented by a rectangular curve.

$$\left. \begin{aligned} -l_1 < x < 0: & \quad \eta + \gamma f_1 = \eta_1 = -\chi_1^2 \\ 0 < x < l_2: & \quad \eta + \gamma f_2 = \eta_2 = -\chi_2^2 \end{aligned} \right\} \quad (44.3)$$

In the first interval we obtain a solution

$$-l_1 < x < 0: \quad u = Ae^{\chi_1 x} + Be^{-\chi_1 x} \quad (44.4)$$

while in the next interval, the solution is

$$0 < x < l_2: \quad u = Ce^{\chi_2 x} + De^{-\chi_2 x} \quad (44.5)$$

Furthermore, the entire solution must fit the form given by Floquet's theorem [Eq. (42.5)]. Choosing one of Floquet's exponentials, as in Eq. (42.6), we write

$$u = A(x)e^{\mu x} \quad (44.6)$$

where $A(x)$ has period d . This means, for instance,

$$u(x) = e^{\mu d} u(x - d) \quad (44.7)$$

Using this relation, we write the solution in the second l_1 interval

$$l_2 < x < l_1 + l_2 = d: \quad u(x) = Ae^{\mu d} e^{\chi_1(x-d)} + Be^{\mu d} e^{-\chi_1(x-d)} \quad (44.8)$$

The problem is to find the A , B , C , and D coefficients that satisfy these relations and the continuity condition at the junctions $(0, l_2)$, where u and $\frac{\partial u}{\partial x}$ must join smoothly. This yields four relations:

$$\left. \begin{aligned} x = 0: & \quad u: & A + B &= C + D \\ & \frac{\partial u}{\partial x}: & A\chi_1 - B\chi_1 &= C\chi_2 - D\chi_2 \\ x = l_2: & u: & Ae^{\mu d - \chi_1 l_2} + Be^{\mu d + \chi_1 l_2} &= Ce^{\chi_2 l_2} + De^{-\chi_2 l_2} \\ & \frac{\partial u}{\partial x}: & A\chi_1 e^{\mu d - \chi_1 l_2} - B\chi_1 e^{\mu d + \chi_1 l_2} &= C\chi_2 e^{\chi_2 l_2} - D\chi_2 e^{-\chi_2 l_2} \end{aligned} \right\} \quad (44.9)$$

This set of four linear homogeneous equations in A , B , $-C$, and $-D$ can be solved only if its determinant is zero.

$$\Delta = \begin{vmatrix} 1 & 1 & 1 & 1 \\ \chi_1 & -\chi_1 & \chi_2 & -\chi_2 \\ e^{\mu d - \chi_1 l_1} & e^{\mu d + \chi_1 l_1} & e^{\chi_2 l_2} & e^{-\chi_2 l_2} \\ \chi_1 e^{\mu d - \chi_1 l_1} & -\chi_1 e^{\mu d + \chi_1 l_1} & \chi_2 e^{\chi_2 l_2} & -\chi_2 e^{-\chi_2 l_2} \end{vmatrix} = 0 \quad (44.10)$$

It is a matter of elementary computations to expand this determinant and to obtain the equation

$$Y^2 - 2Y \left[\cosh \chi_1 l_1 \cosh \chi_2 l_2 + \frac{1}{2} \left(\frac{\chi_1}{\chi_2} + \frac{\chi_2}{\chi_1} \right) \sinh \chi_1 l_1 \sinh \chi_2 l_2 \right] + 1 = 0 \quad (44.11)$$

where $Y = e^{\mu d}$. This second-order equation has two solutions (Y_1, Y_2) whose product is unity

$$Y_1 = e^{\mu d}, \quad Y_2 = e^{-\mu d}, \quad Y_1 Y_2 = 1$$

and whose sum is twice the bracket in Eq. (44.11).

$$\left. \begin{aligned} Y_1 + Y_2 &= e^{\mu d} + e^{-\mu d} = 2 \cosh \mu d = 2[\cdot \cdot \cdot] \\ \cosh \mu d &= \cosh \chi_1 l_1 \cosh \chi_2 l_2 \\ &\quad + \frac{1}{2} \left(\frac{\chi_1}{\chi_2} + \frac{\chi_2}{\chi_1} \right) \sinh \chi_1 l_1 \sinh \chi_2 l_2 \end{aligned} \right\} \quad (44.12)$$

To check this equation, let us take $\chi_1 = \chi_2$, a continuous line. Then $\mu d = \chi_1 l_1 + \chi_2 l_2$ is the obvious solution.

We shall discuss a whole class of problems of this type in Chap. X and develop a more direct and very powerful method for solving them. Equation (44.12) will appear as a special case of a more general equation. According to whether χ_1 and χ_2 are real or imaginary, some of the hyperbolic sines and cosines may become ordinary sines and cosines. For instance, van der Pol and Strutt consider the problem of equal sections. Let us take

$$l_1 = l_2 = \frac{\pi}{2} \qquad d = \pi$$

in order to obtain the same period π as occurs in Mathieu's equation,

$$\eta_1 = \eta + \gamma \qquad \eta_2 = \eta - \gamma$$

Case 1:

$$\left. \begin{aligned} \eta > \gamma > 0 & \quad \text{both } \eta_1 \text{ and } \eta_2 \text{ positive} \\ \chi_1 l_1 = \frac{1}{2} \pi \chi_1 = \frac{\pi}{2} \sqrt{-\eta_1} = i x_1 \\ \mu = i\mu' = i2\pi a \\ \chi_2 l_2 = \frac{1}{2} \pi \chi_2 = \frac{\pi}{2} \sqrt{-\eta_2} = i x_2 \end{aligned} \right\} \quad (44.13)$$

Equation (44.12) becomes

$$\cos \pi \mu' = \cos x_1 \cos x_2 - \frac{1}{2} \left(\frac{x_1}{x_2} + \frac{x_2}{x_1} \right) \sin x_1 \sin x_2$$

Case 2:

$$\left. \begin{aligned} \eta < \gamma, \quad \gamma > 0 & \quad \eta_1 > 0, \quad \eta_2 < 0 \\ \frac{\pi}{2} \chi_1 = \frac{\pi}{2} \sqrt{-\eta_1} = i x_1 \\ \frac{\pi}{2} \chi_2 = \frac{\pi}{2} \sqrt{-\eta_2} = x_3 \end{aligned} \right\} \quad (44.14)$$

$$\cos \pi \mu' = \cos x_1 \cosh x_3 - \frac{1}{2} \left(\frac{x_1}{x_3} - \frac{x_3}{x_1} \right) \sin x_1 \sinh x_3 \quad (44.15)$$

The problem is to follow the variation of expressions (44.12) or (44.14) and (44.15) in the $\eta\gamma$ plane and to distinguish between the regions giving μ real or complex. Three cases are obtained:

$$\left. \begin{aligned} A. \quad \cosh \mu d &\geq 1 & \mu &= \mu_0 \text{ real} \\ B. \quad -1 &\leq \cosh \mu d \leq 1 & \mu &= i\mu', \quad \mu' \text{ real} \\ C. \quad -1 &\leq \cosh \mu d & \mu &= \mu_0 + i\pi \end{aligned} \right\} \quad (44.16)$$

Case *B* gives waves propagating through the whole line without attenuation, which means *passing bands*, and is represented in Fig. 44.3 by shaded regions. Both cases *A* and *C* yield waves attenuated exponentially either to the right or to the left [$e^{\pm \mu \infty}$ factor in Eq. (44.6)] and differ only in the relative phases of oscillations in successive (l_1, l_2) sections. This means stopping bands and blank regions in Fig. 44.3.

The whole map in Fig. 44.3 is very similar to that in Fig. 42.1 for Mathieu functions, except for some intersections of boundary curves. The boundaries are obtained for $k = 0$ or $\mu = i\pi$. They are denoted by $C_0 C_1 C_2$ and $S_0 S_1 S_2$, these symbols corresponding to the *Ce* and *Se* Mathieu functions, respectively.

The discussion given by van der Pol and Strutt covers all cases

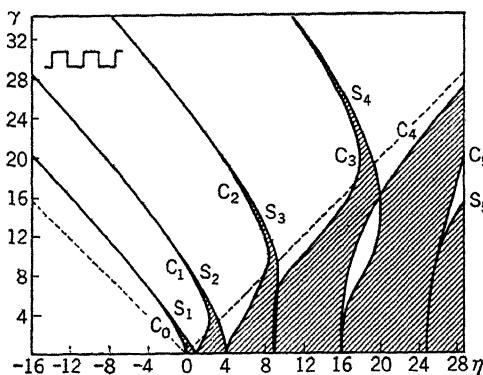


FIG. 44.3.

of (η, γ) variations. Kronig and Penney limited their discussion to the case of $(-\eta_1)$ very large and positive, with l_1 very small, while η_2 is positive and $l_2 \approx d$. The exact conditions assumed are

$$\chi_1 \rightarrow \sqrt{\frac{2c}{l_1 l_2}}, \quad \chi_2 = i\alpha$$

where c is a constant. Then Eq. (44.12) yields

$$\cosh \mu d = \cosh \chi_1 l_1 \cos \alpha l_2 + \frac{\chi_1^2 - \alpha^2}{2\alpha \chi_1} \sinh \chi_1 l_1 \sin \alpha l_2 \quad (44.17)$$

According to these conditions

$$\begin{aligned} \cosh \chi_1 l_1 &\rightarrow 1 \\ \frac{\chi_1^2 - \alpha^2}{2\alpha \chi_1} \sinh \chi_1 l_1 &\rightarrow \frac{l_1}{2\alpha} (\chi_1^2 - \alpha^2) = \frac{c}{\alpha l_2} \end{aligned}$$

and we obtain

$$\cosh \mu d = \cos \alpha d + \frac{c}{\alpha d} \sin \alpha d \quad (44.18)$$

The variation of this expression is represented in Fig. 44.4 where the limits ± 1 corresponding to cases A , B , and C of Eq. (44.16) are clearly seen.

A general formula giving the relation between ν and a was obtained in Sec. 43 by using a result proved by Whittaker under the assumption that the coefficients $\theta_n = \omega^2 C_n$ of the Fourier expansion for the function J constituted an absolutely convergent series. This is not true of the step function used in this section.

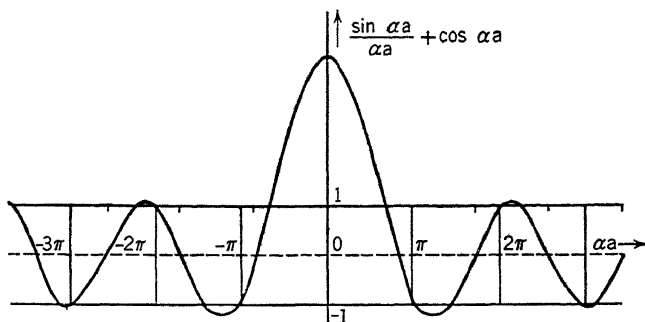


FIG. 44.4.

The function can be expanded in a Fourier series, but the convergence of the series of coefficients is not secured.

Equation (43.16) is

$$\sin^2 \pi a d = \Delta_1(0) \sin^2 \frac{\pi \nu d}{V_0} \quad (43.16)$$

where $\Delta_1(0)$ did not depend upon α or ν , and d was supposed equal to π , while

$$\frac{1}{V_0^2} = \overline{\left(\frac{1}{V^2}\right)} \quad \text{average}$$

In the present section we obtained Eq. (44.14)

$$\cos 2\pi a d = \cos x_1 \cos x_2 - \frac{1}{2} \left(\frac{x_1}{x_2} + \frac{x_2}{x_1} \right) \sin x_1 \sin x_2 \quad (44.14)$$

with

$$x_1 = \frac{\omega d}{2V_1} = \frac{\pi \nu d}{V_1}, \quad x_2 = \frac{\omega d}{2V_2} = \frac{\pi \nu d}{V_2} \quad (d = \pi)$$

This equation can be written

$$\cos 2\pi a d = \cos (x_1 + x_2) + \left(1 - \frac{x_1^2 + x_2^2}{2x_1x_2} \right) \sin x_1 \sin x_2$$

but

$$1 - \frac{x_1^2 + x_2^2}{2x_1x_2} = - \frac{(x_1 - x_2)^2}{2x_1x_2} = \frac{(V_2 - V_1)^2}{2V_1V_2} \quad (44.19)$$

We may replace $\cos 2\varphi$ by $(1 - 2 \sin^2 \varphi)$ and obtain

$$\begin{aligned} \sin^2 \pi a d = \sin^2 \left[\left(\frac{1}{V_1} + \frac{1}{V_2} \right) \frac{\pi \nu d}{2} \right] \\ + \frac{(V_2 - V_1)^2}{4V_1V_2} \sin \frac{\pi \nu d}{V_1} \sin \frac{\pi \nu d}{V_2} \end{aligned} \quad (44.20)$$

which *does not reduce to the general type* (43.16). For *very small perturbations*, however, we may take

$$\frac{1}{V_1} = \frac{1}{V_0} (1 + \epsilon), \quad \frac{1}{V_2} = \frac{1}{V_0} (1 - \epsilon),$$

$$\frac{(V_2 - V_1)^2}{4V_1V_2} = \epsilon^2 + \dots$$

Hence

$$\sin^2 \pi ad = (1 + \epsilon^2 \dots) \sin^2 \left(\frac{\pi v d}{V_0} \right) \quad (44.21)$$

dropping higher powers of ϵ . This equation, rather than the general one, is of the Whittaker type.

This example shows the limitations of the general equations of Sec. 43, which should not be used when the series of the Fourier coefficients θ_n (or C_n) does not converge absolutely.

45. The Self-excited Oscillator

A circuit containing only an inductance and a capacity will oscillate with frequency $1/\sqrt{LC}$, where L is the inductance and C the capacity. The equation of the circuit is

$$L \frac{d^2 Q}{dt^2} + \frac{Q}{C} = 0 \quad (45.1)$$

If we vary the capacity periodically, the equation may be written in the form of Mathieu's equation.

$$\frac{1}{C} = A + B \cos \omega_1 t \quad (45.2)$$

where A and B are constants and ω_1 the frequency with which we vary the capacity. Equation (45.1) becomes

$$\frac{d^2 Q}{dt^2} + \frac{1}{L} (A + B \cos \omega_1 t) Q = 0 \quad (45.3)$$

If we set

$$Q = u \quad \text{and} \quad \xi = \frac{1}{2} \omega_1 t$$

then Eq. (45.3) is

$$\frac{\partial^2 u}{\partial \xi^2} + (\eta + \gamma \cos 2\xi) u = 0 \quad (45.4)$$

where

$$\eta = \frac{4A}{L\omega_1^2}, \quad \gamma = \frac{4B}{L\omega_1^2} \quad (45.5)$$

Equation (45.4) is Mathieu's equation and has solutions of the type (42.5)

$$u = D_1 A(\xi) e^{\mu \xi} + D_2 A(-\xi) e^{-\mu \xi} \quad (45.6)$$

where the exponent μ is pure imaginary in the shaded regions of Fig. 42.1 and real or complex in the blank regions. This means

$$\left. \begin{array}{ll} \text{Shaded regions:} & \mu = i\beta \text{ stable oscillations of constant average amplitude} \\ \text{Blank regions:} & \mu \text{ real or complex; } \textit{unstable} \text{ oscillations, one term increasing to infinity, the other term decreasing to zero} \end{array} \right\} \quad (45.7)$$

Thus the map in Fig. 42.1 completely describes the situation. If the variation of capacity were stepwise instead of sinusoidal, we should use the map in Fig. 44.3.

An example of an electric circuit with variation of the capacity is shown in Fig. 45.1. This device was proposed in the last



FIG. 45.1.

century as a self-excited high-frequency oscillator. A self-excited oscillator can also be built with a periodic variation of the self-inductance, such as a standard alternator with stator and rotor windings connected in series, and a fixed capacity. Self-excitation means unstable conditions, which yield oscillations of increasing amplitude: the amplitude would start from zero and continue to increase, finally reaching a constant value when some of the nonlinear terms omitted in the equation become sufficiently important; such terms might be sparks in the condenser or finite power of the engine turning the condenser.

In the *electric examples*, L and C are *always positive*, which means

$$\eta > \gamma > 0 \quad (45.8)$$

i.e., case 1 in Sec. 44 [Eq. (44.14)], which corresponds to region III in Figs. 42.1, 44.1, and 44.4, where stability is the rule and instability the exception. Looking at Fig. 42.1, we notice that

the main region of instability (or *self-excitation*) is the V-shaped region near $\eta = 1$, which means, according to Eqs. (45.2) and (45.5),

$$\left. \begin{aligned} \frac{4A}{L\omega_1^2} &= \frac{1}{LC_0\omega_1^2} = 1 \frac{\omega_0^2}{\omega_1^2} \approx 1 \\ A &= \frac{1}{C_0}, \quad \omega_0^2 = \frac{1}{LC_0}, \quad 2\omega_0 \approx \omega_1 \end{aligned} \right\} \quad (45.9)$$

ω_0 is the frequency of the circuit (L, C_0) and should be nearly equal to $1/2$ the frequency of excitation ω_1 .

This result was found by an elementary discussion, relating to the case of small excitation ($B \ll A$) before the complete theory was developed. The situation is easier to explain if the

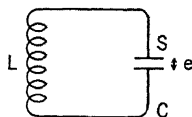


FIG. 45.2.

condenser is supposed to be plane (Fig. 45.2) with a varying distance e between the plates.

$$\frac{1}{C} = \frac{4\pi}{S} e \quad (45.10)$$

If the electric surface density on the condenser plates is $\sigma = Q/S$, then the force acting per unit area is $2\pi\sigma^2$, and the total force of attraction between the plates is

$$f = S2\pi\sigma^2 = \frac{2\pi}{S} Q^2 = \frac{Q^2}{2Ce} \quad (45.11)$$

The variation of capacity is obtained by varying the distance.

$$\left. \begin{aligned} e &= e_0 + b \cos \omega_1 t \\ \frac{1}{C} &= \frac{1}{C_0} + \frac{4\pi b}{S} \cos \omega_1 t \end{aligned} \right\} \quad (45.12)$$

For small perturbation, small b , the oscillations in the circuit will have a frequency very close to the proper frequency ω_0 , and hence Q^2 will vary with twice this frequency.

$$\left. \begin{aligned} Q &= A \cos (\omega_0 t + \varphi) \\ Q^2 &= A^2 \cos^2 (\omega_0 t + \varphi) = \frac{1}{2} A^2 [1 + \cos 2(\omega_0 t + \varphi)] \end{aligned} \right\} \quad (45.13)$$

The work done per oscillation by the engine operating the system is

$$\begin{aligned} W &= \oint f \, de \approx \frac{-b\omega_1 A^2}{2Ce_0} \oint \cos^2(\omega_0 t + \varphi) \sin \omega_1 t \, dt \\ &= \frac{-b\omega_1 A^2}{4Ce_0} \oint \cos 2(\omega_0 t + \varphi) \sin \omega_1 t \, dt \quad (45.14) \end{aligned}$$

This is zero unless $\omega_1 = 2\omega_0$ and may be positive or negative according to the phase φ . When the work done is positive, oscillations in the circuit increase in amplitude (instability and self-excitation). This is a crude explanation of the physical

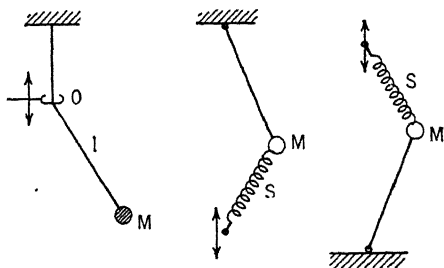


FIG. 45.3.

meaning of condition (45.9). The physical explanation for excitation near the points

$$\eta = 4, 9, \dots, n^2$$

corresponding to

$$\frac{4A}{L_1\omega_1^2} = \frac{4}{LC_0\omega_1^2} = \frac{4\omega_0^2}{\omega_1^2} = n^2; \quad 2\omega_0 = n\omega_1 \quad (45.15)$$

is not so elementary and must be found in the harmonic content of oscillations in the circuit with varying capacity.

Mechanical oscillators can also be taken as examples, and some models are shown in Fig. 45.3. Here the mass M (replacing L) is always positive, but the A and B coefficients in the restoring force [replacing $1/C$, Eq. (45.2)] may become negative, as for a reversed pendulum or for a spring pushing the mass away instead of pulling it toward its equilibrium position. This means that the whole map of Fig. 42.1 or 44.3 can be used. Region II is very interesting: an unstable pendulum can be made stable by a periodic perturbation of appropriate frequency. Some

other examples follow: the movement of a direct or reversed pendulum whose support is moved up and down with a frequency ω_1 ; oscillations along a string whose tension is varied periodically.

More detailed explanations and equations may be found in the paper by van der Pol and Strutt already quoted.

46. Free Electrons in Metals

According to wave mechanics, the motion of electrons in a potential field (x, y, z) is obtained from the solution of the Schrodinger wave equation:

$$\nabla^2\psi + k^2(E - P)\psi = 0, \quad k^2 = \frac{8\pi^2m}{h^2} \quad (46.1)$$

where h is Planck's constant. In a crystal lattice the potential P results from the positive charge of the ions located at the lattice points and from the equal negative charge of the electron cloud distributed among the ions. All this yields a periodic P function with the same periods d_1 , d_2 , and d_3 as the crystal lattice itself, and hence Eq. (46.1) reduces to a type similar to the one studied in Sec. 34, where we had

$$\nabla^2\psi + H(\mathbf{r})\psi = 0 \quad (34.1)$$

$$H = \frac{\omega^2}{V^2} = \omega^2 F \quad (34.2)$$

while here the correspondence is

$$H = k^2[E - P(\mathbf{r})] \quad (46.2)$$

E is the energy of the electrons and $P(\mathbf{r})$ the periodic potential.

The whole theory of Sec. 34 applies directly to the problem of free electrons and yields the usual rules about zone structure. As a matter of fact, the *zone structure* is completely *independent of the special physical meaning* of the waves considered, and it must be the same for *elastic, electromagnetic, and Schrodinger electronic waves*. Some authors did not pay enough attention to this very general result and based their definitions of the zones on different criteria for different waves, thus obtaining discrepancies that are not consistent with the mathematical nature of the problem. As we emphasized in Secs. 26 and 31, another error to be avoided is an oversimplification of the problem, such as an assumption

$$P(\mathbf{r}) = P_1(x) + P_2(y) + P_3(z) \quad (46.3)$$

This is an academic problem of no actual interest whatever, since no crystal lattice is known that would give a field of the type (46.3), even as a first rough approximation. Such a structure would mean a disappearance of most Bragg spots in X rays except for the few reflections from planes parallel to coordinate planes, a circumstance never realized in any known crystal. Actual physical problems yield periodic potentials P with their complete set of A_{m_1, m_2, m_3} coefficients in the triple Fourier expansion, which means a complete set of lattice planes with indices m_1 , m_2 , and m_3 .

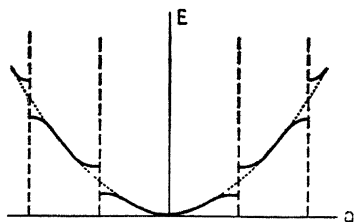


FIG. 46.1a.

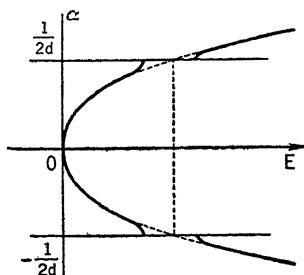


FIG. 46.1b.

The problem of a small perturbation for usual waves was discussed in Sec. 35 under the assumption

$$F = \frac{1}{V^2} = \frac{1}{V_0^2} + \epsilon f = C_{000} + \epsilon f, \quad H = \frac{\omega^2}{V_0^2} + \epsilon \omega^2 f$$

In the electronic problem we may assume the average potential inside the crystal to be zero, since this merely means a special choice of the zero energy level, and we state

$$\left. \begin{aligned} P &= \epsilon \varphi \\ H &= k^2(E - \epsilon \varphi) \end{aligned} \right\} \quad (46.4)$$

Writing Planck's relation

$$E = h\nu \quad (46.5)$$

we define a frequency for the electron waves. The difference between the two problems results from a comparison of the equations

$$\text{Elastic waves } \frac{\omega^2}{V_0^2} = \omega^2 C_{000} = h k^2 \nu \text{ electron waves} \quad (46.6)$$

The V-shaped curves obtained for unperturbed elastic waves with

$$\frac{\omega}{V_0} = 2\pi|a| \quad (46.7)$$

now result in parabolic curves for unperturbed electron waves with

$$\hbar k^2 v = 4\pi^2 |a|^2 \quad (46.8)$$

and curves relating to a small perturbation differ from this parabola only in discontinuities on the boundaries of the zones, as shown in Fig. 46.1.

The whole problem, however, is more complicated than the simple scheme developed here, since the periodic potential P is not known a priori. This potential, as we said before, results from the distribution of both ions and electrons. The average electron density can be computed once the ψ functions of Eq. (46.1) have been obtained. The potential P is deduced from the electron density, and it must check with the P function upon which the whole computation was initially based. This is a typical problem of the self-consistent field, as Hartree calls it, and the solution of Eq. (46.1) represents only one step in a more complicated problem.

Selected References

- MATHIEU, P.: "Cours de Mathématique physique," p. 122, 1873.
 HUMBERT, J.: "Fonctions de Lamé et Mathieu," G. Villars, Paris, 1926.
 WHITTAKER and WATSON: "Modern Analysis," p. 404, Cambridge, 1920.
 STRUTT, M. J. O.: *Ann. Physik*, **84**, 485 (1927); **85**, 129 (1928); **86**, 319 (1928).
 VAN DER POL, B., and M. J. O. STRUTT: *Phil. Mag.*, **5**, 18 (1928).
 BRILLOUIN, L.: "Quantenstatistik," p. 271, Springer, Berlin, 1931.
 KRONIG, R. DE L., and W. G. PENNEY: *Proc. Roy. Soc. (London)*, **130**, 499 (1931).
 STRATTON, J. A.: "Electromagnetic Theory," pp. 52-200, McGraw-Hill, New York, 1941.
 BRILLOUIN, L.: *Electrical Communications*, April, 1938.
 STRUTT, M. J. O.: *Math. Ann.*, **101**, 559 (1929).
 MORSE, P. M.: *Phys. Rev.*, **35**, 1310 (1930).

CHAPTER IX

MATRICES AND THE PROPAGATION OF WAVES ALONG AN ELECTRIC LINE

47. General Remarks

In the historical summary given in the first chapters, it was explained how the theory of wave propagation first started with the discussion of waves along a discontinuous string. Then followed the theory of waves in a continuous medium, and we emphasized the importance of some of Lord Kelvin's remarks on waves in a discontinuous structure and the existence of a cutoff frequency. Up to Kelvin's time only one type of wave had been discussed, *viz.*, elastic waves. Later electromagnetic waves and, still later, electron waves in wave mechanics were discovered, and the properties first obtained for elastic waves were immediately translated for these new waves. For instance, Lagrange's theory of how to pass from the discontinuous string to the limit of a continuous string was used by Pupin in his discussion of loaded telephonic cables. The deep discussion of Kelvin, related in Sec. 2, led him to imagine a new model for an optical medium. A similar mechanical model was built by Vincent and proved to have the properties of a mechanical band-pass filter. This model was translated into an electrical circuit by Campbell and was the point of departure for his invention of electric filters, of which he gave a number of important applications.

Hence, for scientists of the last century, it was common knowledge that the special nature of the waves did not matter and that the same general properties could be found for any type of waves. The general relations among the various types of waves seem to have been forgotten for some time. Physicists developed the theory of electromagnetic waves for optics and X rays. Then theoreticians discussed very carefully the properties of electron waves (wave mechanics) in crystals and too often did not pay attention to the fact that a great part of the work had already been done in the theory of X-ray propagation in crystals.

On the other hand, electrical engineers did a wonderful analysis on the theory of propagation of waves along lines, cables, filters, etc., but omitted to notice that many important facts had already been discovered by theoretical physicists (see Secs. 13 and 14). More recently, practical acoustics was revived, mostly by electrical engineers, who were especially well trained in electric-circuit theory and found it easier to translate mechanical problems into the equivalent electric circuits before discussion. These scientists at last rediscovered the similarity of all problems of vibration and wave propagation, but they did just the opposite of what their ancestors had done. Pupin and Campbell started from mechanical models to discuss electric lines and filters. A modern engineer, wishing to discuss wave propagation along a train of railway cars, translates the problem into an electrical one (an impulse propagating along a filter) and then translates the answer back into mechanical terms.

This explains why we want to include a general discussion of wave propagation along electric lines and filters in this book. Many modern theoretical physicists have hardly heard of these problems and do not realize the very great advance in the theory by this engineer's art. Engineers, on the other hand, have a tendency to imagine that any wave problem can be reduced to a problem in electric lines, and this is not entirely true. We have already discussed in Chap. V the importance and the limitations of the concept of *characteristic* or *surge impedance*. This concept is fundamental for one-dimensional structures such as mechanical or electric lines and filters. Its generalization for three dimensions is not so easy, and we noted that the definition of energy flow, exemplified by the Poynting vector for electromagnetic waves or similar definitions in wave mechanics, is better adapted to the three- or four-dimensional problems.

Recent developments in wave mechanics point to the importance of matrix calculus and its very close connection with a number of problems of wave propagation. It is very interesting to note that electrical engineers have independently come to the same conclusion. Matrix theory is now commonly used in the discussion of problems of waves in electric filters. Furthermore, these problems seem to represent the only classical example where some special matrices, of great importance for the theory of electron spin, appear for practical purposes, and we shall try

in the next chapter to show the connection between the theory of electric filters and the Pauli-Dirac wave theory of the spinning electron.

The discussion in this chapter will be on the propagation of waves along electric lines. Let us, once for all, give the correlation between the electric quantities and the mechanical quantities arising in similar mechanical problems. This can be done in different ways, but the most direct translation is obtained by the use of the following glossary:

<i>Electrical</i>	<i>Mechanical</i>
Electric charge	Displacement or coordinate
Electric current	Velocity
Electric current, time derivative	Acceleration
Self-inductance	Mass
Mutual inductance	No direct equivalent; it appears, however, in problems with constraints, where generalized Lagrange coordinates are used
Magnetic energy	Kinetic energy
Electric energy	Potential energy
Capacity C	Elastic coefficient = $1/C$
Voltage	Tension

We shall discuss in this chapter the propagation of electric disturbances along electric lines. This can be translated into a problem of mechanical disturbances propagating along a periodic mechanical line structure. Electric lines are schematically represented in Fig. 51.1 or 55.1, for instance. The same general scheme may be just as well interpreted in mechanical terms. A square box with four terminals is supposed to represent a certain electric circuit. Let us imagine the box to contain a given mechanical device. We have two terminals on the left for input current and voltage and the two others on the right for output current and voltage when we think of the box as containing an electric circuit. With a mechanical structure, we have only one connection on each side (a rod or a string going from one box to the next one), but we need two quantities to specify the connection: the *velocity* of the string motion (analogous to electric current) and the elastic tension along the string (analogous to the voltage). Hence the correlation is complete, and the glossary will help in translating from one problem to the other.

As noted earlier, there are, however, some cases where

difficulties are encountered: electrical mutual inductances between circuit elements have no direct counterpart in mechanical problems. However, a general expression for the kinetic energy was introduced long ago in analytical mechanics and has the same form as Eq. (48.2) of the next section. Such expressions are obtained in mechanical problems with constraints when generalized Lagrange coordinates must be used instead of the usual position coordinates of the mass points.

Other limitations refer to the range of possible values for some quantities. We have already pointed out in Secs. 11 and 12 the fact that capacities are always positive in electrical theory, while elastic coefficients may be either positive or negative in mechanics. Thus, if the general scheme and theory are common to all problems of wave propagation, there are restricting conditions or practical limitations for each separate class of problems, which should always be kept in mind.

The first sections of this chapter are devoted to essential definitions and to the systematic introduction of the matrix notation, visualizing elementary matrix computations by their equivalent circuit connections. Then we shall go on to a discussion of the role of characteristic impedance of the line and the propagation of waves. Finally, in the next chapter we shall let the elements of the line become infinitesimal and thus will be enabled to draw some analogies between the propagation of the waves and the quantum mechanical problem of electron spin.

The electric lines that we shall discuss are to consist of identical circuits connected together. Each circuit is an electric circuit, which may be as complicated as desired with the following restrictions. First, all circuit elements must be linear; *i.e.*, we allow resistances, self-inductances, mutual inductances, and capacities. Rectifiers, coils containing iron, and other non-linear elements are excluded. Negative resistances are allowed if care is exercised in their use. They will appear in the mathematical equations, and, if they are to be used in an experiment, one must be sure that the linear portion of the characteristic curve is used. The second restriction is that no sources of current are to appear in the circuit. All electromotive forces will appear as external parts.

We shall represent the circuits composing the line as rectangles with pairs of terminals, as shown in Fig. 47.1. At present, we

shall place no restriction on the number of terminals. The electromotive forces are to be applied to these terminals, which we number. Each electromotive force E_k will then give rise to a current i_k , both of the same frequency. For the present we shall choose the signs of E_k and i_k so that the product $R.P. E_k i_k^*$ is the power furnished to the circuit. We shall assume that the circuit is a simple circuit; *i.e.*, that it does not contain two or more circuits completely separate from one another. Further, we shall assume that the number of terminals is reduced to a minimum. For instance, if two or more pairs of terminals are

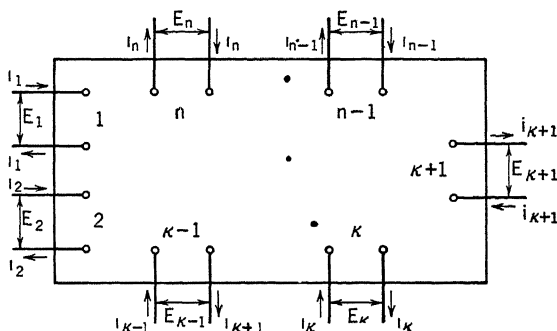


FIG. 47.1.

in series, they may be replaced by a single pair of terminals with an applied electromotive force equal to the sum of the separate electromotive forces.

48. Expressions for Energy

We shall assume that the minimum number of pairs of terminals is n . The k th pair of terminals will have an electromotive force E_k . The meaning of $E_k = 0$ is just that the k th pair of terminals is short-circuited. Each electromotive force E_k will furnish charge q_k to the pair of terminals across which it is connected. The q_k will form a complete set of independent variables for the system if there are n branches to the circuit. If there are more than n , then it is necessary to introduce more terminals with an applied electromotive force of zero at these additional terminals. We shall assume that this has been done.

There will be two types of energy present in the general circuit: electrostatic due to capacities, magnetic due to self-inductances and mutual inductances, and, in addition, dissipation of energy

due to resistances. If we let Q_α be the charge on condenser α , then the electrostatic energy is given by

$$E_e = \frac{1}{2} \sum_{\alpha} \frac{Q_\alpha^2}{C_\alpha} \quad (48.1)$$

where C_α is the capacity of condenser α . Similarly, if \dot{Q}_α is the current in inductance α , the magnetic energy is

$$E_m = \frac{1}{2} \sum_{\alpha} L_\alpha \dot{Q}_\alpha^2 + \sum_{\alpha > \beta} M_{\alpha\beta} \dot{Q}_\alpha \dot{Q}_\beta = \frac{1}{2} \sum_{\alpha\beta} M_{\alpha\beta} \dot{Q}_\alpha \dot{Q}_\beta \quad (48.2)$$

where

$L_\alpha = M_{\alpha\alpha}$ = self-inductance of inductance α

$M_{\alpha\beta} = M_{\beta\alpha}$ = mutual inductance of inductances α and β
when $\alpha \neq \beta$

Finally, if \dot{Q}_α is the current flowing through resistance α , the dissipated energy per unit time is

$$\phi = \sum_{\alpha} R_\alpha \dot{Q}_\alpha^2 \quad (48.3)$$

where R_α is the resistance of resistance α .

Now all of our circuit elements are to be linear and it follows that we may express the Q_α as a linear sum of the q_k defined in the first paragraph of this section; i.e.,

$$Q_\alpha = \sum_k a_{\alpha k} q_k \quad (48.4)$$

where the $a_{\alpha k}$ are constant coefficients. We may substitute Eq. (48.4) into Eqs. (48.1), (48.2), and (48.3) to obtain first

$$E_e = \frac{1}{2} \sum_{\alpha} \frac{1}{C_\alpha} \left(\sum_k a_{\alpha k} q_k \sum_j a_{\alpha j} q_j \right) = \frac{1}{2} \sum_{kj} c_{kj} q_k q_j \quad (48.5)$$

where

$$c_{kj} = \sum_{\alpha} \frac{a_{\alpha k} a_{\alpha j}}{C_\alpha} = c_{jk} \quad (48.6)$$

Next

$$E_m = \frac{1}{2} \sum_{\alpha\beta} M_{\alpha\beta} \sum_j a_{\alpha j} \dot{q}_j \sum_k a_{\beta k} \dot{q}_k = \frac{1}{2} \sum_{jk} m_{jk} \dot{q}_j \dot{q}_k \quad (48.7)$$

where

$$m_{jk} = \sum_{\alpha\beta} M_{\alpha\beta} a_{\alpha j} a_{\beta k} = m_{kj} \quad (48.8)$$

Finally,

$$\phi = \sum_{\alpha} R_{\alpha} \sum_j a_{\alpha j} \dot{q}_j \sum_k a_{\alpha k} \dot{q}_k = \sum_{jk} r_{jk} \dot{q}_j \dot{q}_k \quad (48.9)$$

where

$$r_{jk} = \sum_{\alpha} R_{\alpha} a_{\alpha j} a_{\alpha k} = r_{kj} \quad (48.10)$$

The coefficients c_{kj} , m_{kj} , and r_{kj} may be computed in terms of the circuit elements; they will be homogeneous in the C_{α} , $M_{\alpha\beta}$, and R_{α} , respectively. Furthermore, they are all symmetrical in k and j .

The general circuit equations may be written

$$E_k = \sum_i (m_{ki} \ddot{q}_i + r_{ki} \dot{q}_i + c_{ki} q_i) \quad (48.11)$$

where E_k is the electromotive force across the terminals k . The work furnished to the circuit in time dt is

$$\begin{aligned} dW &= \sum_k E_k \dot{q}_k dt = \sum_k \sum_i (m_{ki} \ddot{q}_i + r_{ki} \dot{q}_i + c_{ki} q_i) \dot{q}_k dt \\ &= \sum_{ik} (m_{ki} \dot{q}_k d\dot{q}_i + r_{ki} \dot{q}_i \dot{q}_k dt + c_{ki} q_i dq_k) \\ &= dE_m + dE_e + \phi dt \end{aligned}$$

This means that we have not omitted any energy from consideration.

We shall assume that all the applied electromotive forces have the same frequency ω . Then

$$\dot{q}_k = \dot{q}_k = I_k e^{j\omega t}, \quad E_k = V_k e^{j\omega t} \quad (j = \sqrt{-1}) \quad (48.12)$$

and Eq. (48.11) may be written

$$V_k = \sum_i \zeta_{ki} I_i \quad (48.13)$$

where

$$\zeta_{ki} = m_{ki}j\omega + r_{ki} + \frac{c_{ki}}{j\omega} = \zeta_{ik} \quad (48.14)$$

from Eqs. (48.6), (48.8), and (48.10).

49. Definition of a Four-terminal and Equations for Its Circuit

A four-terminal is a special case of the circuits we have discussed in the previous sections. It is a circuit for which

$$E_k = 0 \quad k \neq 1, 2 \quad (49.1)$$

For a four-terminal, Eq. (48.13) becomes

$$\left. \begin{aligned} \sum_i \zeta_{1i} I_i &= V_1 \\ \sum_i \zeta_{2i} I_i &= V_2 \\ \sum_i \zeta_{ki} I_i &= 0 \quad k = 3, 4, \dots, n \end{aligned} \right\} \quad (49.2)$$

where n is the number of branches. We may solve the equations (49.2) since the number of equations is equal to the number of independent variables I_i .

$$\left. \begin{aligned} I_1 &= \chi_{11} V_1 + \chi_{12} V_2 \\ I_2 &= \chi_{21} V_1 + \chi_{22} V_2 \\ I_k &= \chi_{k1} V_1 + \chi_{k2} V_2 \quad k = 3, 4, \dots, n \end{aligned} \right\} \quad (49.3)$$

It can be shown that $\chi_{ki} = \chi_{ik}$ follows from $\zeta_{ki} = \zeta_{ik}$. The first two relations are the only ones of interest. Solving them for V_1 and V_2 gives

$$\left. \begin{aligned} V_1 &= z_{11} I_1 + z_{12} I_2 \\ V_2 &= z_{21} I_1 + z_{22} I_2 \end{aligned} \right\} \quad z_{21} = z_{12} \quad (49.4)$$

where the z_{ij} are constants of the circuit.

So far we have taken the V_i and I_i to have signs so that the product $E_i \dot{q}_i$ is the power furnished by the electromotive force. A different convention will be more useful in the discussion of electric lines since we shall regard the line as composed of a row of four-terminals each with a current flowing in one direction along one set of connections and in the opposite direction along

the other set. This is shown in Fig. 49.1. If we consider four-terminal n , we have a potential difference E_1 at the left end and E_2 at the right end. The current I'_1 at the left end flows into the four-terminal, and I'_2 flows out of the right end. Then I'_1 and I'_2 bear the following relation to I_1 and I_2 :

$$I'_1 = I_1, \quad I'_2 = -I_2 \quad (49.5)$$

Then Eq. (49.4) becomes on dropping the prime from I'_1 and I'_2

$$\left. \begin{aligned} V_1 &= z_{11}I_1 - z_{12}I_2 \\ V_2 &= z_{12}I_1 - z_{22}I_2 \end{aligned} \right\} \quad (49.6)$$

There is one further change in the equations to be made for convenience in discussing line problems. In general, we are

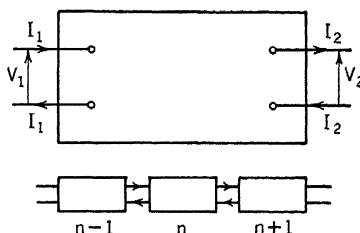


FIG. 49.1.

interested in comparing conditions at one end of the four-terminal with conditions at the other end, rather than currents with electromotive forces. This means that we should express V_1 and I_1 in terms of V_2 and I_2 . Rearranging Eqs. (49.6) gives

$$\left. \begin{aligned} V_1 &= a_1V_2 + bI_2 \\ I_1 &= cV_2 + a_2I_2 \end{aligned} \right\} \quad (49.7)$$

where

$$\left. \begin{aligned} a_1 &= \frac{z_{11}}{z_{12}}, & a_2 &= \frac{z_{22}}{z_{12}} \\ b &= \frac{z_{11}z_{22} - z_{12}^2}{z_{12}}, & c &= \frac{1}{z_{12}} \end{aligned} \right\} \quad (49.8)$$

on taking account of the fact that $z_{12} = z_{21}$. The four constants are not independent, for it is obvious that

$$a_1a_2 - bc = 1 \quad (49.9)$$

50. Matrix Notation for a Four-terminal

We shall find it useful to regard the current and voltage at the exit of a four-terminal as two quantities that are transformed

by a matrix to give the current and voltage at the entrance. The matrix will, of course, be the matrix of the coefficients in Eq. (49.7). From this point of view, it is convenient to change our notation as follows:

$$\left. \begin{aligned} \text{Electromotive force at entrance} &= V_1 = x_1 \\ \text{Electromotive force at exit} &= V_2 = x'_1 \\ \text{Current at entrance} &= I_1 = x_2 \\ \text{Current at exit} &= I_2 = x'_2 \end{aligned} \right\} \quad (50.1)$$

Then Eq. (49.7) may be written

$$\left. \begin{aligned} x_1 &= a_{11}x'_1 + a_{12}x'_2 \\ x_2 &= a_{21}x'_1 + a_{22}x'_2 \end{aligned} \right\} \quad (50.2)$$

where

$$\left. \begin{aligned} a_{11} &= a_1, & a_{12} &= b \\ a_{21} &= c, & a_{22} &= a_2 \end{aligned} \right\} \quad (50.3)$$

Then we may call the matrix (a_{ij}) the matrix of the four-terminal. It depends only on the constants of the four-terminal circuit and is thus characteristic of the four-terminal.

It follows immediately from Eqs. (49.9) and (49.3) that the determinant of the matrix (a_{ij}) is unity.

$$|a_{ij}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} = a_1a_2 - bc = 1 \quad (50.4)$$

Equation (50.2) may be solved for x'_1 and x'_2 in terms of x_1 and x_2 .

$$\left. \begin{aligned} x'_1 &= b_{11}x_1 + b_{12}x_2 \\ x'_2 &= b_{21}x_1 + b_{22}x_2 \end{aligned} \right\} \quad (50.5)$$

where

$$\left. \begin{aligned} b_{11} &= \frac{a_{22}}{|a|} = a_{22}, & b_{12} &= -\frac{a_{12}}{|a|} = -a_{12} \\ b_{21} &= -\frac{a_{21}}{|a|} = -a_{21}, & b_{22} &= \frac{a_{11}}{|a|} = a_{11} \end{aligned} \right\} \quad (50.6)$$

It is evident that the determinant of the matrix (b_{ij}) is unity. The matrix (b_{ij}) is the inverse of the matrix (a_{ij}) , and we may write Eqs. (50.2) and (50.5) symbolically as follows:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}; \quad \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = B \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \quad B = A^{-1} \quad (50.7)$$

where A and B stand for the matrices (a_{ij}) and (b_{ij}) , respectively.

Thus we see that there are two matrices, one the inverse of the other, that are of significance for a given four-terminal. One of these, A , would be used if we thought of a wave propagating through the four-terminal from right to left, and the other would be used for a wave propagating from left to right.

It should be noted that the elements of the matrices A and B are not all pure numbers; some have dimensions:

$a_{11}, a_{22}, b_{11}, b_{22}$	pure numbers
a_{12}, b_{12}	impedance
a_{21}, b_{21}	admittance (reciprocal of impedance)

51. Combination of Two Four-terminals; Multiplication of Matrices

In this and the following two sections, we shall illustrate various rules from the theory of matrices by means of four-terminals. First, we consider the effect of connecting two four-

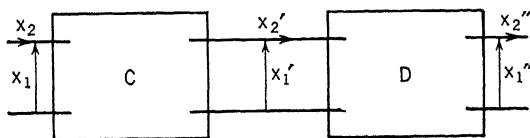


FIG. 51.1.

terminals in cascade, as in Fig. 51.1. These two four-terminals together constitute a composite four-terminal. We wish to find the matrix of the resultant four-terminal in terms of the matrices of the two component four-terminals.

If the matrix (b_{ij}) of the previous section is denoted by (c_{ij}) for the left-hand four-terminal and by (d_{ij}) for the right-hand four-terminal, then

$$\left. \begin{aligned} x''_i &= \sum_j d_{ij} x'_j = \sum_j d_{ij} \left(\sum_k c_{jk} x_k \right) \\ &= \sum_k \left(\sum_j d_{ij} c_{jk} \right) x_k = \sum_k f_{ik} x_k \end{aligned} \right\} \quad (51.1)$$

where the double-primed letters represent conditions at the exit (right end) of the right-hand four-terminal, the single-primed letters conditions between the two four-terminals, and the unprimed letters conditions at the entrance of the left-hand four-terminal.

Now

$$f_{ik} = \sum_j d_{ij} c_{jk} \quad \text{or} \quad F = D \cdot C \quad (51.2)$$

are the elements of a matrix (f_{ik}) that is the product of the matrices (d_{ij}) and (c_{ij}) . The order in which the matrices appear makes a difference in the product except in certain cases. We shall have occasion to note some of these cases in later sections. In general, however,

$$\sum_j c_{ij} d_{jk} = C \cdot D \neq D \cdot C = \sum_j d_{ij} c_{jk} \quad (51.3)$$

In the previous section we noted that A and B were inverse matrices. It is readily verified by direct calculation that

$$A \cdot B = B \cdot A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \delta \quad (51.4)$$

δ is called the unit matrix since it is evident that

$$A \cdot \delta = \delta \cdot A = A \quad (51.5)$$

for any arbitrary matrix A . Moreover, it is also easily shown that the determinant of the product P of any two matrices R and S is the product of the determinants of R and S .

$$|P| = |R \cdot S| = |R| \cdot |S| \quad (51.6)$$

52. Inverse and Reversed Four-terminals and Transformations

If two four-terminals, connected in cascade, produce no change in the electromotive force or in the current, then we say that one four-terminal is the *inverse* of the other; *i.e.*, if A is the matrix of the first and B the matrix of the second,

$$A \cdot B = \delta = B \cdot A \quad (52.1)$$

In this case, evidently, the order of the four-terminals with respect to the direction of propagation is immaterial.

Now making use of the fact that A and B each have determinant one and Eq. (50.6), we see that

$$\left. \begin{aligned} a_{11} &= b_{22} & a_{12} &= -b_{12} \\ a_{21} &= -b_{21} & a_{22} &= b_{11} \end{aligned} \right\} \quad (52.2)$$

Evidently to construct the four-terminal inverse to that with matrix A will require self and mutual inductances, capacities,

and resistances that are the negatives of those occurring in four-terminal A . This is not always impossible, but often quite difficult to achieve experimentally.

By a *reversed four-terminal*, we mean one in which the sense of the current is changed and the entrance and exit are interchanged. This is, it should be noted, not the same as changing the direction of propagation of waves. The latter involves only interchanging the entrance and exit.

We shall denote the transformation of a four-terminal by B and the corresponding one for the reversed four-terminal by R . Then the direct four-terminal gives

$$\begin{aligned}x'_1 &= b_{11}x_1 + b_{12}x_2 \\x'_2 &= b_{21}x_1 + b_{22}x_2\end{aligned}$$

Interchanging the entrance and exit requires us to use the matrix (a_{ij}) inverse to (b_{ij}) .

$$\begin{aligned}x'_1 &= a_{11}x_1 + a_{12}x_2 \\x'_2 &= a_{21}x_1 + a_{22}x_2\end{aligned}$$

and changing the sign of the currents (x_2 and x'_2) yields

$$\begin{aligned}x'_1 &= a_{11}x_1 - a_{12}x_2 \\x'_2 &= -a_{21}x_1 + a_{22}x_2\end{aligned}$$

Therefore, the matrix $R = (r_{ij})$ is given by

$$\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & -a_{12} \\ -a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} b_{22} & b_{12} \\ b_{21} & b_{11} \end{pmatrix} \quad (52.3)$$

on making use of $|a| = 1$ and Eq. (50.6). Evidently,

$$|r_{ij}| = |a_{ij}| = |b_{ij}| = 1 \quad (52.4)$$

The elements of the matrix of the reversed four-terminal are equal to elements of the matrix of the direct four-terminal, so the construction of the reverse of a given four-terminal will not involve negative resistances except as they occur in the direct four-terminal.

We call a four-terminal *reversible* if it is identical with its reverse; *i.e.*, a four-terminal is reversible if

$$R = B \quad (52.5)$$

which yields

$$b_{11} = b_{22} \quad (52.6)$$

Since the determinant of B is unity,

$$b_{11} = b_{22} = \sqrt{1 + b_{12}b_{21}} \quad (52.7)$$

so that the matrix of a reversible four-terminal is

$$B = \begin{pmatrix} \sqrt{1 + b_{12}b_{21}} & b_{12} \\ b_{21} & \sqrt{1 + b_{12}b_{21}} \end{pmatrix} \quad (52.8)$$

53. Four-terminal Matrices and the Group C_2

The remarks that we have made in the previous sections are sufficient to show that the matrices arising in the theory of four-terminals form a group of two-rowed complex matrices of determinant one. The conditions that must be fulfilled in order for a set of matrices to form a group of the above type are

1. The matrices must be two-rowed complex matrices with determinant one (see Sec. 50).

2. Each matrix must possess an inverse matrix that is an element of the group. Evidently each four-terminal must have an inverse, since the number of equations involved is equal to the number of independent variables, and hence the matrix is non-singular and possesses an inverse (see Sec. 50).

3. Two matrices multiplied together must give a matrix in the group (see Sec. 51).

4. There is a unit matrix, *i.e.*, δ (see Sec. 51).

The group composed of matrices of four-terminals is at least a subgroup of the group C_2 . The group C_2 is well known to mathematicians and is an integral part of the theory of electron spin and relativistic quantum mechanics. We shall find these matrices appearing in a similar fashion in the next chapter on the propagation of waves along lines composed of infinitesimal circuits.

Matrices have been introduced in the preceding sections in connection with the electrical problem of four-terminals. It is important to compare these definitions with the standard geometrical definitions given by mathematicians: In a plane, two coordinates x_1 and x_2 define a vector \mathbf{P} , while x'_1 and x'_2 yield \mathbf{P}' . The fundamental linear relations (50.7) show that the matrix A transforms any arbitrary \mathbf{P}' vector into another \mathbf{P} , while the inverse matrix B brings the transformed \mathbf{P} back to \mathbf{P}' .

This geometrical representation is perfectly adequate when

the x 's are real numbers. In our electrical example, we always deal with complex numbers, and Fig. 53.1 must be considered only as a geometrical visualization of the matrix properties.

Multiplication of matrices was explained in Sec. 51 as representing the connection of two four-terminals in cascade. Looking at Eqs. (51.1) and (51.2), we notice that they represent two successive transformations of the vector \mathbf{P} first from \mathbf{P} to \mathbf{P}' by matrix C and then from \mathbf{P}' to \mathbf{P}'' by matrix D .

The reversed four-terminal defined in Sec. 52 is typical of the electrical problem. Its geometrical counterpart was never considered in the theory of matrices. It can, however, be stated this way: We take \mathbf{P}_1 to be the reflection in the x_1 axis of \mathbf{P}' ,

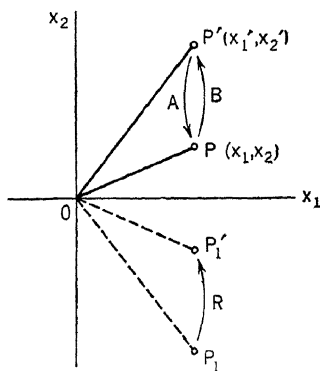


FIG. 53.1.

and \mathbf{P}'_1 the reflection in the x_1 axis of \mathbf{P} . The R matrix (52.3) represents the transformation from \mathbf{P}_1 to \mathbf{P}'_1 . A reversible four-terminal is represented by a matrix built in such a way that R is identical with B where R transforms \mathbf{P}_1 into \mathbf{P}'_1 .

The next step in the geometrical description of matrices is to look for the *axes* of the matrix, which are defined by the condition that \mathbf{P} and \mathbf{P}' lie in the same direction.

$$\left. \begin{aligned} \mathbf{P}' &= \xi \mathbf{P} \\ x'_1 &= \xi x_1 = b_{11}x_1 + b_{12}x_2 \\ x'_2 &= \xi x_2 = b_{21}x_1 + b_{22}x_2 \end{aligned} \right\} \quad (53.1)$$

These two linear equations can be solved only if their determinant is zero.

$$\begin{vmatrix} b_{11} - \xi & b_{12} \\ b_{21} & b_{22} - \xi \end{vmatrix} = 0 \quad (53.2)$$

This is known as the equation for the *proper values* of the matrix, and it yields two ξ coefficients, one corresponding to each of the axes of the matrix. The orientations of the axes are obtained when the ξ values of Eq. (53.2) are used in Eq. (53.1). Their slopes are given by the condition

$$S = \frac{x'_2}{x'_1} = \frac{x_2}{x_1}$$

$$S = \frac{b_{21} + b_{22}S}{b_{11} + b_{12}S} \quad (53.3)$$

$$b_{12}S^2 + (b_{11} - b_{22})S - b_{21} = 0 \quad (53.4)$$

We shall soon discover the physical meaning of these quantities and find that they correspond to very important definitions in the theory of electric four-terminals.

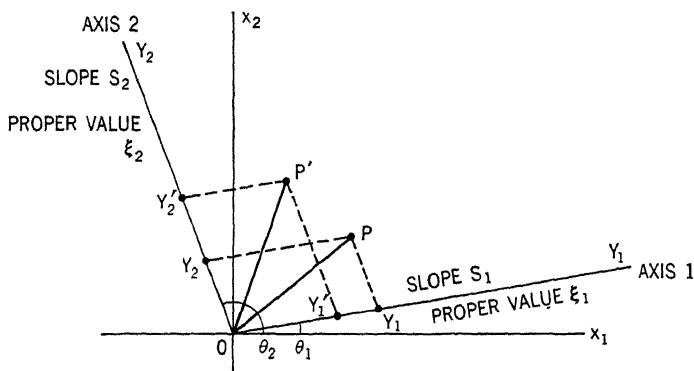


FIG. 53.2.

Once the axes of the matrix and the proper values have been determined, the matrix transformation acquires a simple geometrical meaning. A vector P is decomposed into its components Y_1 and Y_2 along the axes. Then each component is multiplied by the corresponding proper value, yielding

$$Y'_1 = \xi_1 Y_1, \quad Y'_2 = \xi_2 Y_2 \quad (53.5)$$

which represent the components of the transformed vector P' , as shown in Fig. 53.2.

This was just explained for the matrix (b_{ij}) , of which (a_{ij}) is the inverse matrix. Matrix (a_{ij}) has the same axes as (b_{ij}) , but its proper values are ξ_1^{-1} and ξ_2^{-1} , and the (a_{ij}) transformation carries P' into P .

The relation between (Y_1, Y_2) and (x_1, x_2) is easily obtained. Let us state

$$S_1 = \tan \theta_1, \quad S_2 = \tan \theta_2$$

Then

$$\left. \begin{aligned} x_1 &= Y_1 \cos \theta_1 + Y_2 \cos \theta_2 \\ x_2 &= Y_1 \sin \theta_1 + Y_2 \sin \theta_2 \end{aligned} \right\} \quad (53.6)$$

which we can also write

$$\left. \begin{aligned} x_2 &= y_1 + y_2, & y_1 &= Y_1 \sin \theta_1 \\ x_1 &= \frac{1}{S_1} y_1 + \frac{1}{S_2} y_2, & y_2 &= Y_2 \sin \theta_2 \end{aligned} \right\} \quad (53.7)$$

We shall use these last formulas for a comparison with the four-terminal problem.

The geometrical representation gives a very simple explanation of the following theorem, which says that *two matrices C and D whose axes coincide are commutative*, in the sense of Eq. (51.3).

$$C \cdot D = D \cdot C \quad (53.8)$$

This results directly from the fact that after the decomposition along the common Y_1 and Y_2 axes the transformation reduces to usual multiplication.

$$Y'_1 = \xi_C \xi_D Y_1 = \xi_D \xi_C Y_1$$

and the same is true for Y_2 .

54. Surge, Iterative, or Characteristic Impedance of a Four-terminal

In Chap. V we discussed the characteristic impedance of a one-dimensional mechanical lattice in some detail. The characteristic impedance of the lattice was taken equal to the mechanical impedance offered by a single cell with its particles vibrating as if a single wave were propagating through an infinite lattice. We shall define the impedance of an electrical line composed of four-terminals in a similar fashion.

We take the impedance connected at the exit of a four-terminal to be z' ; i.e.,

$$x'_1 = z' x'_2 \quad z' = \text{output impedance} \quad (54.1)$$

Substitution of this relation in Eq. (50.2) yields

$$\left. \begin{aligned} x_1 &= (a_{11}z' + a_{12})x'_2 = (b_{22}z' - b_{12})x'_2 \\ x_2 &= (a_{21}z' + a_{22})x'_2 = (-b_{21}z' + b_{11})x'_2 \end{aligned} \right\} \quad (54.2)$$

upon making use of Eqs. (50.4) and (50.6). Now a four-terminal with its exit shunted by an impedance z' has an impedance z at the entrance given by

$$z = \frac{x_1}{x_2} = \frac{b_{22}z' - b_{12}}{-b_{21}z' + b_{11}} = \text{input impedance} \quad (54.3)$$

A similar calculation may be made for the reversed four-terminal. We take ζ and ζ' for the input and output impedances, respectively (Fig. 54.1).

$$\zeta = -\frac{b_{11}\zeta' - b_{12}}{b_{21}\zeta' - b_{22}} \quad (54.4)$$

from Eq. (52.3) for the matrix of a reversed four-terminal.

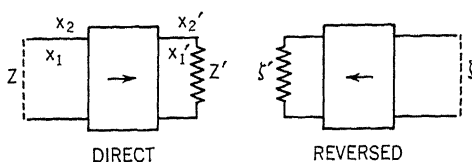


FIG. 54.1.

The four-terminal will have two iterative, surge, or characteristic impedances, obtained by the condition that

$$z = z' \quad \text{or} \quad \zeta = \zeta' \quad (54.5)$$

These two conditions are equivalent. The two roots of

$$b_{21}z^2 + (b_{22} - b_{11})z - b_{12} = 0 \quad (54.6)$$

are z and $-\zeta$. Equation (54.6) is obtained from Eq. (54.3). The analogous equation obtained from Eq. (54.4) has roots $-\zeta$ and ζ .

Comparing Eq. (54.3) with Eq. (53.3), we note the relation

$$z = \frac{1}{S_1}, \quad \zeta = -\frac{1}{S_2}$$

The characteristic impedance is the inverse of the slope of the axis of the matrix (Fig. 53.2). Equation (54.6) is thus similar to Eq. (53.4). The choice of z and $-\zeta$ as characteristic imped-

ances would yield two positive numbers in our geometrical representation instead of a positive and a negative one. It is just a matter of conventions and is connected with the definition of the reversed four-terminal.

The two impedances z and ζ are called the *surge* or *characteristic* impedances in the direct and in the reverse senses, respectively. They are often denoted by k_1 and k_2 . A *reversible four-terminal* is characterized by $b_{11} = b_{22}$ and hence

$$k_1 = k_2, \quad z = \zeta \quad (54.7)$$

The axes in Fig. 53.2 are symmetrical with respect to the x_1 and x_2 coordinate axes in this case.

One of the characteristic impedances corresponds to the situation obtained in a row of identical four-terminals when a pure wave is propagating from left to right along the row. The

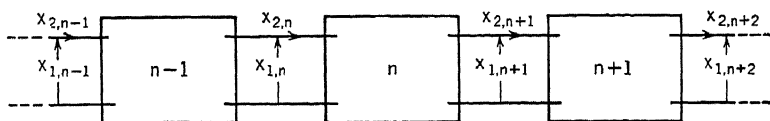


FIG. 55.1.

other characteristic impedance corresponds to a wave propagating from right to left. This will be explained in Sec. 55.

55. Propagation along a Line of Four-terminals

We start by assuming an infinite number of four-terminals connected in cascade, as in Fig. 55.1. Later we shall see how to terminate the line without disturbing the propagation of waves along it.

A single wave that propagates along this infinite line is characterized by the fact that the electromotive force and current are multiplied by the same complex factor ξ as the wave passes from four-terminal n to four-terminal $(n + 1)$.

$$\left. \begin{aligned} x_{1,n+1} &= \xi x_{1,n} \\ x_{2,n+1} &= \xi x_{2,n} \end{aligned} \right\} \quad (55.1)$$

Equation (55.1) together with Eq. (50.5) gives us two linear homogeneous equations in $x_{1,n}$ and $x_{2,n}$:

$$\left. \begin{aligned} \xi x_{1,n} &= b_{11}x_{1,n} + b_{12}x_{2,n} \\ \xi x_{2,n} &= b_{21}x_{1,n} + b_{22}x_{2,n} \end{aligned} \right\} \quad (55.2)$$

which can be solved only if the determinant of the coefficients vanishes.

$$\begin{vmatrix} b_{11} - \xi & b_{12} \\ b_{21} & b_{22} - \xi \end{vmatrix} = \xi^2 - (b_{11} + b_{22})\xi + 1 = 0 \quad (55.3)$$

where we make use of the fact that the determinant $|b_{ij}| = 1$. Thus the complex factor ξ is determined by the constants of the four-terminal circuit.

The equation (55.3) is well known in the theory of matrices. The solutions for ξ are the *proper values* of the matrix (b_{ij}) , or the diagonal elements if the matrix is reduced to diagonal form, as explained in Eq. (53.2). The two solutions are

$$\xi_{1,2} = \frac{1}{2}(b_{11} + b_{22}) \pm \sqrt{\frac{1}{4}(b_{11} + b_{22})^2 - 1} \quad (55.4)$$

Let us take

$$\xi = e^\gamma = e^{\alpha + i\beta}$$

Then, since $\xi_1 \xi_2 = 1$ from Eq. (55.4), we may write

$$\begin{aligned} \xi_1 &= e^\gamma = e^{\alpha + i\beta} & \frac{\xi_1 + \xi_2}{2} &= \cosh \gamma = \frac{1}{2}(b_{11} + b_{22}) \\ \xi_2 &= e^{-\gamma} = e^{-\alpha - i\beta} \end{aligned} \quad (55.5)$$

If $\alpha = 0$, then $|\xi_1| = |\xi_2| = 1$ and

$$\xi_1 = e^{-i\beta}, \quad \xi_2 = e^{i\beta} \quad (55.6)$$

In this case one obtains propagation of waves without attenuation, and the two solutions given in Eq. (55.6) correspond to propagation in opposite directions: ξ_1 gives propagation to the right and ξ_2 to the left.

If $\alpha \neq 0$, attenuation is present. α is the attenuation constant and β the change in phase per four-terminal. When $\alpha > 0$, ξ_1 and ξ_2 give propagation to the right and left, respectively, as before.

Let us compute the ratio of electromotive force to current at the entrance to four-terminal n . From Eq. (55.2), we obtain

$$\frac{x_{1,n}}{x_{2,n}} = \frac{b_{12}}{\xi - b_{11}} = \frac{\xi - b_{22}}{b_{21}} \quad (55.7)$$

which is a constant complex number that we may call z . There are two values of z , z_1 and z_2 , corresponding to ξ_1 and ξ_2 . A simple calculation shows that these two solutions are the two

characteristic impedances corresponding to the directions of propagation associated with ξ_1 and ξ_2 . Thus in the notation of the last paragraph of the previous section

$$z_1 = k_1, \quad z_2 = -k_2 = -\zeta$$

We may ask how to terminate the line at four-terminal n in such a way as to avoid reflection of waves coming from the left. This is done by using an output impedance k_1 on the right side of four-terminal n . In the same way we terminate a line on the left side and prevent reflection for waves coming from the right by using an impedance k_2 on the left of the last four-terminal.

An arbitrary wave propagating along a line of four-terminals may be split up into a sum of two simple waves traveling in opposite directions. Evidently, from Eq. (55.1), the matrix of each four-terminal is diagonal for the simple waves. This procedure of splitting the vibration up into two waves is analogous to taking the principal axes for a transformation in matrix theory. Thus we may take the current to be y_1 for propagation to the right and y_2 for propagation to the left. Then the total electromotive force will be given by the sum of the electromotive forces $z_1 y_1$ and $z_2 y_2$ for the two currents, and the total current will be the sum of y_1 and y_2 . We add a subscript n to the y 's to indicate the four-terminal under consideration. Then if we take $n = 0$ for the first four-terminal,

$$x_{1,0} = z_1 y_{1,0} + z_2 y_{2,0}, \quad x_{2,0} = y_{1,0} + y_{2,0} \quad (55.8)$$

These equations are identical with those for the reduction of the matrix to its axis, given in Eq. (53.7). Equation (55.8) holds for any n since we assume a stable condition in the line. Equation (55.8) may also be written in the form

$$\left. \begin{aligned} y_{1,0} &= \frac{x_{1,0} - z_2 x_{2,0}}{z_1 - z_2} \\ y_{2,0} &= \frac{-x_{1,0} + z_1 x_{2,0}}{z_1 - z_2} \end{aligned} \right\} \quad (55.9)$$

These equations will be of use shortly.

Now to obtain the current and electromotive force at four-terminal n , we merely note that

$$y_{1,n} = \xi_1^n y_{1,0}, \quad y_{2,n} = \xi_2^n y_{2,0} \quad (55.10)$$

where

$$\left. \begin{aligned} \xi_1^n &= e^{-n\gamma} = e^{-n(\alpha + j\beta)} \\ \xi_2^n &= e^{n\gamma} = e^{n(\alpha + j\beta)} \end{aligned} \right\} \quad (55.11)$$

We have taken α positive when the propagation to the right is represented by ξ_1 . In this case ξ_1^n becomes negligible for very large n while ξ_2^n is large.

If we have an arbitrary impedance $\zeta_0 = -x_{1,0}/x_{2,0}$ placed at the beginning of the line, the impedance at the n th four-terminal will be, for very large n ,

$$\zeta_n = -\frac{x_{1,n}}{x_{2,n}} \approx -\frac{z_2 \xi_2^n y_{2,0}}{\xi_2^n y_{2,0}} = -z_2 \quad (55.12)$$

since from Eqs. (55.8) and (55.10) the exact solution is

$$\left. \begin{aligned} x_{1,n} &= z_1 \xi_1^n y_{1,0} + z_2 \xi_2^n y_{2,0} \\ x_{2,n} &= \xi_1^n y_{1,0} + \xi_2^n y_{2,0} \end{aligned} \right\} \quad (55.13)$$

and ξ_1^n is negligible for very large n . This means that, except for the first few four-terminals, the impedance for very large n does not depend on the impedance at the end. This impedance is one of the characteristic impedances of the four-terminals that make up the line; the other one z_1 is obtained by reversing the condition.

Substitution of Eqs. (55.9) and (55.11) into Eq. (55.13) yields the relations

$$\left. \begin{aligned} x_{1,n} &= \frac{1}{z + \zeta} [(\zeta e^{n\gamma} + z e^{-n\gamma})x_{1,0} - 2\zeta z (\sinh n\gamma)x_{2,0}] \\ x_{2,n} &= \frac{1}{z + \zeta} [-2(\sinh n\gamma)x_{1,0} + (z e^{n\gamma} + \zeta e^{-n\gamma})x_{2,0}] \end{aligned} \right\} \quad (55.14)$$

where we have set

$$z_1 = z, \quad z_2 = -\zeta \quad (55.15)$$

Equation (55.14) is called the canonical form of the line equations. They contain only three constants: the two characteristic impedances of the four-terminal and the propagation constant γ .

56. Application of the Theory to a Reversible Four-terminal

Equation (54.6) gives us the characteristic impedance of a four-terminal in terms of the constants of its circuit. A reversible four-terminal is characterized by the condition $b_{11} = b_{22}$; hence,

for a reversible four-terminal we have

$$z = \zeta = \sqrt{\frac{b_{12}}{b_{21}}} \quad (56.1)$$

This value is the geometric mean of the impedances of the four-terminal with open circuit and with short circuit. For the open circuit $z' = \infty$, and $z = -(b_{22}/b_{21})$, while for the short circuit $z' = 0$ and $z = -(b_{12}/b_{11})$ [Eq. (54.3)]. The impedances

$$z_{I_1} = \sqrt{\frac{b_{12}}{b_{21}} \frac{b_{22}}{b_{11}}} \quad \text{and} \quad z_{I_2} = \sqrt{\frac{b_{12}}{b_{21}} \frac{b_{11}}{b_{22}}} \quad (56.2)$$

are called the *image impedances*. They are the geometric mean of the impedance z on open circuit and short circuit for the four-terminal and the reversed four-terminal, respectively. They

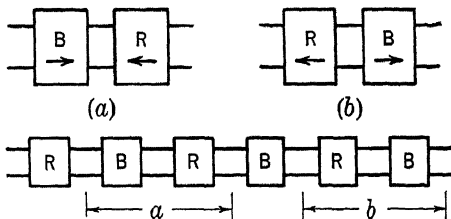


FIG. 56.1.

are the same and equal to the characteristic impedance in the case of a reversible four-terminal. This coincidence disappears for other cases.

Equation (55.14) becomes for a reversible four-terminal ($z = \zeta$)

$$\left. \begin{aligned} x_{1,n} &= (\cosh n\gamma)x_{1,0} - (z \sinh n\gamma)x_{2,0} \\ x_{2,n} &= \left(-\frac{1}{z} \sinh n\gamma\right)x_{1,0} + (\cosh n\gamma)x_{2,0} \end{aligned} \right\} \quad (56.3)$$

The image impedances (56.2) are not directly connected with the properties of the row of similar four-terminals. This is better shown by proving that they represent the surge impedances of a row of symmetrical reversible four-terminals obtained by joining a given four-terminal B to its reverse R . Figure 56.1 shows a row of alternate R and B four-terminals. It can be considered either as a row of cells in the order (B,R) or as a row of cells in the order (R,B) .

Let us take the first case and write down the corresponding

matrix [Eq. (51.2)], which is simply the product of R and B matrices in this order:

a. (B, R) four-terminal

$$\text{Matrix: } R \cdot B = \begin{pmatrix} r_{11}b_{11} + r_{12}b_{21} & r_{11}b_{12} + r_{12}b_{22} \\ r_{21}b_{11} + r_{22}b_{21} & r_{21}b_{12} + r_{22}b_{22} \end{pmatrix} \quad (56.4)$$

When the r_{ik} coefficients from Eq. (52.3) are substituted,

$$R \cdot B = \begin{pmatrix} b_{11}b_{22} + b_{12}b_{21} & 2b_{12}b_{22} \\ 2b_{11}b_{21} & b_{11}b_{22} + b_{12}b_{21} \end{pmatrix} \quad (56.5)$$

This represents a reversible four-terminal since the two diagonal elements are equal [Eq. (52.6)]. It has only *one* surge impedance, which, according to Eq. (56.1), is given by the square root of the ratio of the nondiagonal elements, and this is just z_{I_1} , the first image impedance. The second combination yields z_{I_2} . Electrical engineers have frequently paid too much attention to these image impedances, which correspond to no essential property of the B matrix itself.

57. Passing Bands and Attenuation in a Line of Four-terminals

Whether the waves propagating along a line of four-terminals are attenuated or not is determined by α [Eq. (55.5)]. If $\alpha = 0$, then there is no attenuation and the waves will be passed by the line. This condition yields

$$\xi_1 = \cos \beta \pm j \sin \beta \quad (57.1)$$

If we set $b = \frac{1}{2}(b_{11} + b_{22})$, then Eq. (55.4) becomes

$$\xi_1 = b \pm \sqrt{b^2 - 1} = b \pm j \sqrt{1 - b^2} \quad (57.2)$$

Comparison of Eqs. (57.1) and (57.2) shows that

$$\cos \beta = b \text{ real, } \quad -1 < b < 1 \quad (57.3)$$

This is the general *condition for propagation* of waves *without attenuation*. It can be obtained also from (55.5) since $\alpha = 0$ leaves us with

$$\cosh \gamma = \cos \beta = b = \frac{1}{2}(b_{11} + b_{22})$$

If our four-terminal circuits contain no resistance, the ξ_{ik} of Eq. (48.14) and the χ_{ik} and the z_{ik} of Eqs. (49.3) and (49.4) are all pure imaginary. It follows that a_1 and a_2 of Eq. (49.8) are real and b and c of Eq. (49.8) are pure imaginary. Hence, the

matrix elements for a four-terminal without resistance are

$$\text{with } \left. \begin{array}{l} b_{11}, b_{12}, b_{21}, b_{22} \\ b_{12} = j\beta_{12} \quad \text{and} \quad b_{21} = j\beta_{21} \end{array} \right\} \quad (57.4)$$

where b_{11} , b_{22} , β_{12} , and β_{21} are all real. Thus for this case the conditions that propagation without attenuation occur are

$$\left. \begin{array}{l} b_{11}, b_{22} \text{ real} \\ -2 \leq b_{11} + b_{22} \leq 2 \end{array} \right\} \quad (57.5)$$

If resistances are present, the first of these conditions is replaced by Eq. (57.3), which says that the sum of b_{11} and b_{22} shall be real, i.e., that b_{11} and b_{22} be complex conjugates. This is equivalent to saying that, in general, we may have propagation without attenuation if $|b_{11} + b_{22}| \leq 2$ and either the circuit contains zero resistance, or positive and negative resistances occur so that the net resistance is zero ($b_{11} = b_{22}^*$).

For a reversible four-terminal, the characteristic impedance is given by Eq. (55.7).

$$z = \frac{\xi - b_{22}}{b_{21}} = \frac{1}{b_{21}} \left[\frac{1}{2} (b_{11} - b_{22}) \pm j \sqrt{1 - \frac{1}{4} (b_{11} + b_{22})^2} \right] \quad (57.6)$$

on using Eq. (57.2). For a reversible four-terminal $b_{11} - b_{22} = 0$ and hence

$$z = \pm \frac{j \sqrt{1 - \frac{1}{4} (b_{11} + b_{22})^2}}{b_{21}} \quad (57.7)$$

If Eq. (57.3) is satisfied, both numerator and denominator of Eq. (57.7) are pure imaginary, which means that z is real. Thus z is real in a passing band and pure imaginary for other frequencies for a reversible four-terminal. The statement cannot, of course, be extended to nonreversible four-terminals.

58. Reflected Waves in a Line Terminated by an Impedance ζ_0

Let us assume z and ζ are the characteristic impedances of a line of four-terminals, and ζ_0 is the impedance at the left end, terminated by four-terminal 0. Then

$$-\zeta_0 = \frac{x_{1,0}}{x_{2,0}} \quad (58.1)$$

A wave propagated to the left will be partly reflected. With the notation of Sec. 55,

$$\begin{aligned} \rho = \text{coefficient of reflection} &= \frac{y_{1,0}}{y_{2,0}} \\ &= \frac{-x_{1,0} + z_2 x_{2,0}}{x_{1,0} - z_1 x_{2,0}} = \frac{z_2 + \zeta_0}{-\zeta_0 - z_1} = \frac{\zeta - \zeta_0}{\zeta_0 + z} \end{aligned} \quad (58.2)$$

since $z_2 = -\zeta$. Evidently, if $\zeta_0 = \zeta$, there will be no reflection.

From this it follows that, to connect two lines of four-terminals without reflection, the characteristic impedances of the two must be equal. In all other cases there will be at least partial reflection (for frequencies in the passing band of the line that receives waves) and total reflection for waves in the stopping bands of the receiving line. The lack of reflection when the characteristic impedances are equal is very closely connected with the theorem explained at the end of Sec. 53 [Eq. (53.8)]. Four-terminals having the same characteristic impedances are represented by matrices whose axes have the same slope and coincide. Such matrices are commutative. This corresponds to the possibility of reversing the order of the four-terminals without changing the properties of the line. Now if the four-terminals C and D taken in either order, (C,D) or (D,C) give the same result, it certainly means that there is no reflection at their junction.

All these properties of four-terminal lines and their characteristic impedances represent a systematic generalization of the simple problems discussed in the first chapters.

59. A Continuous Line Loaded with Two-terminals

In this section we shall consider a line loaded with two-terminals as an example of the power of the matrix method.

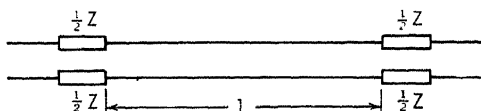


FIG. 59.1.

Such a line is shown in Fig. 59.1. The line consists of impedances $\frac{1}{2}z$ spaced a distance l from one another on each of the upper and lower wires of the line. This is very similar to the line loaded with uniformly spaced equal self-inductances discussed in Sec. 11. Here, however, we do not specify the elements contributing to the impedance $\frac{1}{2}z$ except to assume that there is no resistance; *i.e.*, z is pure imaginary.

$$z = jz', \quad z' \text{ real} \quad (59.1)$$

The impedance z' of any arbitrary two-terminal can be shown to vary with the frequency according to the general law

$$z' = \frac{(\omega^2 - \omega_2^2)(\omega^2 - \omega_4^2) \cdots}{(\omega^2 - \omega_1^2)(\omega^2 - \omega_3^2) \cdots} L\omega \quad (59.2)$$

where $\omega = 2\pi\nu$ and $\omega_1 < \omega_2 < \omega_3 < \omega_4 \dots$. Equation (59.2) is valid as long as the two-terminal contains only a finite number of circuit elements and no portions of a continuous line. A typical curve of z' as a function of ω is plotted in Fig. 59.2. The

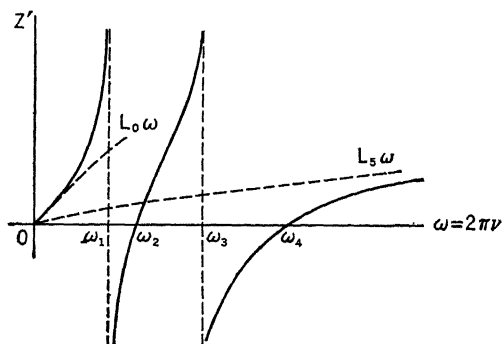


FIG. 59.2.

points ω_1 and ω_3 correspond to points of antiresonance, and ω_2 and ω_4 correspond to points of resonance. The number of such points can be increased at will by properly choosing the arrangement of the two-terminals, since this determines the number of branches in the (Z, ω) curve. An impedance law (59.2) can be obtained with a number of different circuits, some examples of which are shown on Fig. 59.3. In Fig. 59.3a, L_5 is the self-inductance at very high frequencies, and the circuit L_1C_1 has a proper frequency ω_1 , while L_3C_3 has a proper frequency ω_3 .

$$L_1C_1\omega_1^2 = 1 \quad L_3C_3\omega_3^2 = 1$$

Frequencies ω_2 and ω_4 lie between ω_1 and ω_3 and above ω_3 . Another type of circuit is shown on Fig. 59.3 where

$$L_0 = L_1 + L_3 + L_5, \quad L_2C_2\omega_2^2 = 1, \quad L_4C_4\omega_4^2 = 1$$

We shall find different types of passing and stopping bands in the continuous line loaded with two-terminals, some of them being characteristic of the two-terminal and some others of the loaded structure. The frequencies $\omega_1, \omega_2, \omega_3, \omega_4 \dots$ are characteristic of the two-terminals used for loading. In the neigh-

borhood of ω_2 and ω_4 , $z' = 0$ and the line works as if it were not loaded at all; hence we find passing bands. In the neighborhood of ω_1 and ω_3 we find $z' = \infty$, hence stopping bands. These bands are referred to as *two-terminal bands*. In addition, the arrangement of the two-terminals may yield further bands, and these we refer to as *structure bands*. These latter bands depend principally on the distance of separation of the two-terminals.

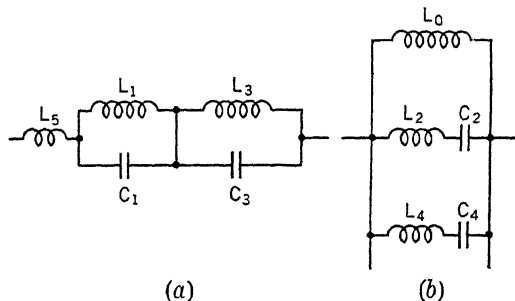


FIG. 59.3.

Now let us apply the theory of matrices to this line to obtain further information on these bands. There will be two parts of the line to be considered: the two-terminals themselves and the lines joining them. The matrix for the lines is obtained from Eq. (55.14) or (56.3). In this case the two values of the characteristic impedance ζ and z are to be set equal:

$$\zeta = z = k \quad (59.3)$$

and we assume zero resistance so that γ , the propagation constant per unit length, is given by

$$\gamma = j\beta \quad (59.4)$$

Then the matrix is obtained from Eq. (56.3), where we substitute l for n , assuming γ to correspond to a unit length of the line.

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} \cosh \gamma l & -k \sinh \gamma l \\ -\frac{1}{k} \sinh \gamma l & \cosh \gamma l \end{pmatrix} \quad (59.5)$$

From inspection, we see that the matrix for the two-terminal itself is

$$\begin{pmatrix} b'_{11} & b'_{12} \\ b'_{21} & b'_{22} \end{pmatrix} = \begin{pmatrix} 1 & -z \\ 0 & 1 \end{pmatrix} \quad (59.6)$$

Then the matrix for a complete section of the line is given by

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} b'_{11} & b'_{12} \\ b'_{21} & b'_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ = \begin{pmatrix} \cosh \gamma l + \frac{z}{k} \sinh \gamma l & -z \cosh \gamma l - k \sinh \gamma l \\ -\frac{1}{k} \sinh \gamma l & \cosh \gamma l \end{pmatrix} \quad (59.7)$$

To investigate the passing bands of our line loaded with two-terminals, we refer to Eq. (57.5). According to this equation, the line will pass frequencies whenever

$$|B_{11} + B_{22}| \leq 2 \quad (59.8)$$

The other conditions in Eq. (57.5) are automatically satisfied because of our assumption that z and γ are pure imaginary. Using Eqs. (59.7) and (59.8), we obtain the condition

$$|\cos B| = \left| \cosh \gamma l + \frac{z}{2k} \sinh \gamma l \right| = \left| \cos \beta l - \frac{z'}{2k} \sin \beta l \right| \leq 1 \quad (59.9)$$

The limits of the passing bands are given by

$$|\cos B| = \left| \cos \beta l - \frac{z'}{2k} \sin \beta l \right| = 1 \quad (59.10)$$

There are two cases to be considered, corresponding to the two types of bands mentioned earlier:

1. *Two-terminal Bands*.—Two-terminal bands contain the point $z' = 0$, since for this case Eq. (59.9) is always satisfied. There will be a certain range of values for z' including the point $z' = 0$ for which Eq. (59.9) is satisfied, and this range comprises the two-terminal passing band.

2. *Structure Bands*.—A structure passing band will have for one of its limits

$$\beta l = \frac{\omega l}{W} = N\pi \quad \text{where } \beta = \frac{\omega}{W}$$

since this gives

$$\sin \beta l = 0, \quad \cos \beta l = \pm 1 \quad (59.11)$$

and condition (59.10) is satisfied. The width of the band can be obtained in the following way: If one limit of the band corresponds to $\cos B_1 = \pm 1$ (when $B_1 = N\pi$), the other limit is found for $\cos B_2 = \mp 1$ and $B_2 = (N+1)\pi$. Let us assume, for instance, $|z'| \gg k$; the second limit is obtained for

$$\begin{aligned}
 \beta l &= N\pi + \epsilon & \epsilon \text{ small} \\
 \cos (N+1)\pi &= -\cos N\pi = \cos \beta l - \frac{z'}{2k} \sin \beta l \\
 &= \cos (N\pi + \epsilon) - \frac{z'}{2k} \sin (N\pi + \epsilon) \\
 &\approx \cos N\pi - \frac{z'}{2k} \cos N\pi \cdot \epsilon
 \end{aligned}$$

Hence

$$\epsilon = \frac{4k}{z'} \quad (59.12)$$

This gives the width of narrow bands, assuming that z' is approximately constant throughout this passing band, and $|z'| \gg k$.

Let us apply this theory to two simple cases: a line loaded with inductances, and a line loaded with capacitances.

a. *A line loaded with inductances.*—In this case

$$z' = L_0 \omega$$

where L_0 is the value of the inductances. $z' = 0$ only for $\omega = 0$. The corresponding two-terminal passing band is the well-known low-frequency band. The structure passing bands are found for high frequencies. One of the band limits is

$$\omega = \frac{N\pi W}{l} \quad (59.13)$$

and the band width is given by

$$\epsilon = \frac{4k}{L_0 \omega} = \frac{4kl}{L_0 N \pi W} = \frac{4}{N \pi} \frac{L_s}{L_0} \quad (59.14)$$

since, if L_s and C_s are the inductance and capacity per section l of the line,

$$k = \sqrt{\frac{L_s}{C_s}}, \quad W = \frac{l}{\sqrt{L_s C_s}} \quad (59.15)$$

Figure 59.4 shows the curve of ω vs.

$$B = \cos^{-1} \left| \cos \beta l - \frac{z'}{2k} \sin \beta l \right|.$$

b. *A line loaded with condensers.*—In this case we obtain

$$z' = -\frac{1}{C_0 \omega}, \quad \epsilon = \frac{4k}{z'} = -4k C_0 \omega, \quad (59.16)$$

where C_0 is the capacity of the condensers and ϵ the width of the

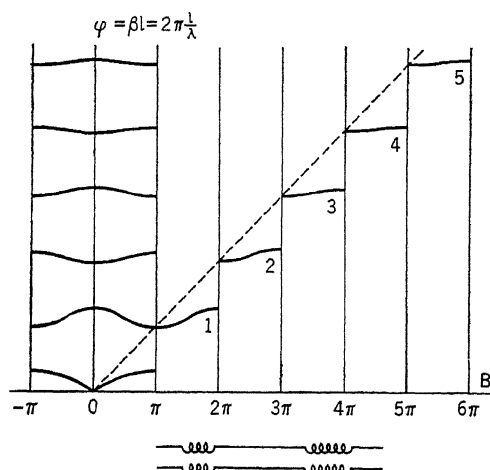


FIG. 59.4.

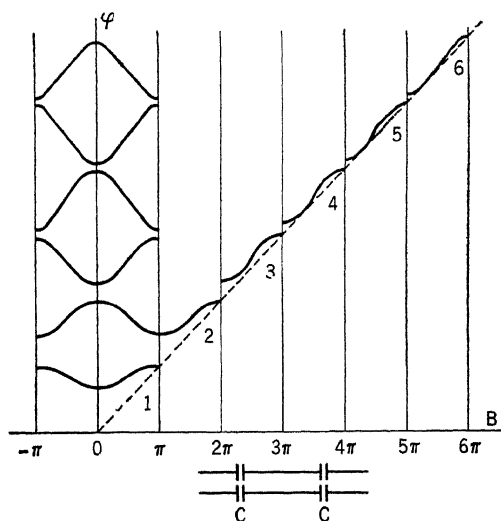


FIG. 59.5.

structure bands. The two-terminal band disappears since $z' = 0$ only for $\omega = \infty$. The upper limits of the structure bands again appear at

$$\omega = \frac{N\pi W}{l} \quad (59.17)$$

Figure 59.5 shows the frequency as a function of B .

60. A Continuous Line Loaded with Four-terminals

In this section we shall treat a problem similar to that of the last section except that we replace the two-terminals with four-terminals. Then the matrix (b'_{ij}) for the four-terminals has the general form given in Eq. (55.14). The matrix (b_{ij}) for the line

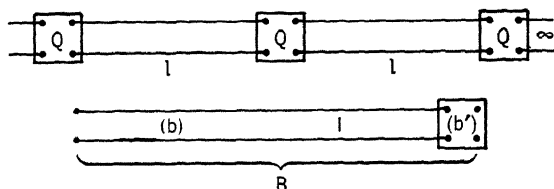


FIG. 60.1.

connecting the four-terminals is given by Eq. (59.5). Multiplying these matrices as we did in the last section, we obtain for one section of the line (see Fig. 60.1)

$$\begin{pmatrix} b'_{11} & b'_{12} \\ b'_{21} & b'_{22} \end{pmatrix} = \begin{pmatrix} \xi e^{\gamma'} + z e^{-\gamma'} & -2z\xi \sinh \gamma' \\ -2 \sinh \gamma' & z e^{\gamma'} + \xi e^{-\gamma'} \end{pmatrix} \frac{1}{z + \xi} \quad (60.1)$$

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} \cosh \gamma l & -k \sinh \gamma l \\ -\frac{1}{k} \sinh \gamma l & \cosh \gamma l \end{pmatrix} \quad (60.2)$$

$$\begin{aligned} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} &= \begin{pmatrix} b'_{11} & b'_{12} \\ b'_{21} & b'_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \frac{1}{z + \xi} \left[\begin{aligned} &\frac{2z\xi}{k} \sinh \gamma' \sinh \gamma l + \cosh \gamma l (z e^{-\gamma'} + \xi e^{\gamma'}) \\ &-2 \sinh \gamma' \cosh \gamma l - \left(\frac{z}{k} e^{-\gamma'} + \frac{\xi}{k} e^{-\gamma'} \right) \sinh \gamma l \\ &-k \sinh \gamma l (z e^{-\gamma'} + \xi e^{\gamma'}) - 2z\xi \sinh \gamma' \cosh \gamma l \\ &2k \sinh \gamma' \sinh \gamma l + \cosh \gamma l (z e^{\gamma'} + \xi e^{-\gamma'}) \end{aligned} \right] \quad (60.3) \end{aligned}$$

where k is the characteristic impedance and γ the propagation

constant of the line net unit length, z and ζ are the characteristic impedances, γ' is the propagation constant of the four-terminal, and l is the distance between four-terminals.

We assume that the four-terminals contain no resistance that gives the condition

$$\left. \begin{array}{ll} b'_{11}, b'_{22} & \text{real} \\ b'_{12}, b'_{21} & \text{pure imaginary} \end{array} \right\} \quad (60.4)$$

In a passing band of the four-terminals, $\gamma' = j\beta'$ is pure imaginary, and this together with condition (60.4) tells us that

$$z = \zeta^* \quad (60.5)$$

In a stopping band of the four-terminals

$$\gamma' = \alpha' \text{ real,} \quad z \text{ and } \zeta \text{ pure imaginary} \quad (60.6)$$

If the four-terminal is reversible, we know that z and ζ are equal and hence real in the passing bands, a result obtained in Eq. (57.7).

Now the condition for a passing band in the line as a whole is given by

$$\begin{aligned} |\cos B| &= \frac{1}{2} |B_{11} + B_{22}| \\ &= \left| \cosh \gamma' \cosh \gamma l + \frac{k + (z\zeta/k)}{z + \zeta} \sinh \gamma' \sinh \gamma l \right| \leq 1 \end{aligned} \quad (60.7)$$

Again we shall find that we can divide the passing bands into two classes: (1) *four-terminal bands* due to the four-terminals, and (2) *structure bands* due to the spacing of the four-terminals.

To obtain the four-terminal bands we must consider frequencies in the passing bands of the four-terminal. Both $\gamma = j\beta$ and $\gamma' = j\beta'$ must be pure imaginary. We set

$$\left. \begin{array}{l} z = z_r + jz_i \\ \zeta = z_r - jz_i \end{array} \right\} \quad (60.8)$$

on using Eq. (60.5). z_r and z_i are the real and imaginary parts, respectively, of z and ζ . This gives on referring to Eq. (60.7)

$$|\cos B| = \left| \cos \beta' \cos \beta l - \frac{k^2 + z_r^2 + z_i^2}{2kz_r} \sin \beta' \sin \beta l \right| \leq 1 \quad (60.9)$$

Passing bands will include points for which

$$\frac{k^2 + z_r^2 + z_i^2}{2kz_r} = \pm 1 \quad \cos B = \cos (\beta' \pm \beta l) \quad (60.10)$$

holds. To obtain further results, we must know the characteristic impedances of the four-terminal as functions of the frequency. If the four-terminal is reversible, $z_r = 0$, and condition (60.10) reduces to

$$z_r = \pm k \quad (60.11)$$

The condition (60.11) means that the loaded line has passing bands containing all frequencies (in the passing bands of the four-terminals) for which the characteristic impedance of the four-terminal equals that of the line (no reflection at their junction). For frequencies in a stopping band of the four-terminal,

$$z = jz', \quad \zeta = j\zeta' \quad (60.12)$$

Equation (60.7) becomes

$$|\cos B| = \left| \cosh \alpha' \cos \beta l + \frac{k^2 - z'\zeta'}{k(z' + \zeta')} \sinh \alpha' \sin \beta l \right| \leq 1 \quad (60.13)$$

In general, Eq. (60.13) is not satisfied. However, there will be a structure passing band around the frequency for which the $\cos B$ is zero. This is given by

$$\cot \beta l = \frac{z'\zeta' - k^2}{k(z' + \zeta')} \tanh \alpha' \quad (60.14)$$

In connection with this problem it is interesting to note that the matrix method greatly simplifies some of the computations discussed in Chap. VIII. For a reversible four-terminal, Eq. (60.7) becomes

$$\left| \cosh \gamma' \cosh \gamma l + \frac{1}{2} \left(\frac{k}{z} + \frac{z}{k} \right) \sinh \gamma' \sinh \gamma l \right| \leq 1 \quad (60.15)$$

since $z = \zeta$. This problem corresponds exactly to the one discussed in Sec. 44, and if we set

$$\left. \begin{aligned} \gamma' &= \chi_1 l_1, & \gamma l &= \chi_2 l_2 \\ k &= \chi_1, & z &= \chi_2 \end{aligned} \right\} \quad (60.16)$$

we obtain Eq. (44.12). The direct computation given in Sec. 44 led to a fourth-order determinant that is equivalent to the determinant of the matrix in Eq. (60.3). The theory developed in this chapter thus appears as an important generalization of the problem of Sec. 44.

Selected References

BRILLOUIN, L.: *Rev. gén. élec.*, **39**, 3-16 (Jan. 4, 1936); **42**, 771-778, (Dec. 18, 1937); **42**, 803-816 (Dec. 23, 1937).

CHAPTER X

CONTINUOUS ELECTRIC LINES

61. Transition from a Line of Four-terminals to a Continuous Line

In the previous sections we discussed lines composed of four-terminals. Each four-terminal was characterized by a matrix that produced finite changes in the current and electromotive force. We may think of these four-terminals as becoming very small so that their matrices produce very small changes and in the limit these changes will be zero. Thus, if dz is an infinitesimal portion of the line, we may take the matrix (b_{ij}) to be

$$(b_{ij}) = \begin{pmatrix} 1 + \epsilon_{11} dz & \epsilon_{12} dz \\ \epsilon_{21} dz & 1 + \epsilon_{22} dz \end{pmatrix} \quad (61.1)$$

which differs very little from the unit matrix. Equation (61.1) may be written

$$(b_{ij}) = e^{(\epsilon_{ij})dz} = 1 + (\epsilon_{ij})dz + \dots \quad (61.2)$$

where we neglect terms in dz higher than the first and

$$(\epsilon_{ij}) = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix}$$

Now the determinant of (b_{ij}) must be equal to one. This condition is fulfilled to the first order in dz by setting

$$\epsilon_{11} = -\epsilon_{22} \quad (61.3)$$

Since our transformation is now an infinitesimal transformation, it follows that

$$\left. \begin{aligned} x'_1 &= b_{11}x_1 + b_{12}x_2 = x_1 + dx_1 = x_1 + (\epsilon_{11}x_1 + \epsilon_{12}x_2)dz \\ x'_2 &= b_{21}x_1 + b_{22}x_2 = x_2 + dx_2 = x_2 + (\epsilon_{21}x_1 + \epsilon_{22}x_2)dz \end{aligned} \right\} \quad (61.4)$$

from which we obtain

$$\left. \begin{aligned} \frac{dx_1}{dz} &= \epsilon_{11}x_1 + \epsilon_{12}x_2 \\ \frac{dx_2}{dz} &= \epsilon_{21}x_1 + \epsilon_{22}x_2 \end{aligned} \right\} \quad \epsilon_{22} = -\epsilon_{11} \quad (61.5)$$

For the special case of a reversible four-terminal, $b_{11} = b_{22}$ or $\epsilon_{11} = \epsilon_{22}$. This result combined with Eq. (61.3) yields

$$\epsilon_{11} = \epsilon_{22} = 0 \quad (61.6)$$

An example is given by a telegraphic cable. Let the cable have series resistance, series self-inductance, shunt conductance, and shunt capacity per unit length given by R , L , G , and C , respectively. Then

$$dV = (R + i\omega L)dzI, \quad dI = (G + i\omega C)dzV$$

which in the usual notation becomes

$$\frac{dx_1}{dz} = \epsilon_{12}x_2, \quad \frac{dx_2}{dz} = \epsilon_{21}x_1 \quad (61.7)$$

where

$$\epsilon_{11} = \epsilon_{22} = 0, \quad \left. \begin{aligned} \epsilon_{12} &= R + i\omega L \\ \epsilon_{21} &= G + i\omega C \end{aligned} \right\} \quad (61.8)$$

The first of Eqs. (61.8) shows that the cable is composed of reversible four-terminals. The four-terminals are, of course, infinitesimal.

We may obtain equations for the propagation of waves along a line composed of nonreversible infinitesimal four-terminals by splitting the waves up into two simple waves propagating in opposite directions, as before. If we let γ be the propagation constant (attenuation and phase shift) per unit length of the line, a single wave propagating in a given direction is represented by

$$\frac{dx_1}{dz} = \gamma x_1, \quad \frac{dx_2}{dz} = \gamma x_2 \quad (61.9)$$

Combining Eq. (61.5) with Eq. (61.9), we obtain

$$\left. \begin{aligned} (\epsilon_{11} - \gamma)x_1 + \epsilon_{12}x_2 &= 0 \\ \epsilon_{21}x_1 - (\gamma + \epsilon_{11})x_2 &= 0 \end{aligned} \right\} \quad (61.10)$$

The condition that Eq. (61.10) be soluble for x_1 and x_2 is that the determinant of the coefficients vanish. The resulting equation

for γ is

$$\gamma^2 - \epsilon_{11}^2 - \epsilon_{12}\epsilon_{21} = 0$$

or

$$\gamma = \pm \sqrt{\epsilon_{11}^2 + \epsilon_{12}\epsilon_{21}} \quad (61.11)$$

The values of γ given by Eq. (61.11) are the proper values of the matrix (ϵ_{ij}) . They are, in general, complex. For the case of reversible four-terminals

$$\gamma = \pm \sqrt{\epsilon_{12}\epsilon_{21}}$$

and

$$\frac{x_1}{x_2} = k = \frac{\epsilon_{12}}{\gamma} = \pm \sqrt{\frac{\epsilon_{12}}{\epsilon_{21}}} \quad (61.12)$$

k is the characteristic impedance of the line and has the same value for the two directions of propagation given by $\pm \gamma$. The characteristic impedances in the general case are different.

$$\left. \begin{matrix} k_1 \\ -k_2 \end{matrix} \right\} = \frac{x_1}{x_2} = \frac{\epsilon_{12}}{\pm \gamma - \epsilon_{11}} = \frac{\pm \gamma + \epsilon_{11}}{\epsilon_{21}} \quad (61.13)$$

62. Examples of Four-terminal Representation of Continuous Lines

A line of infinitesimal four-terminals may be used to represent certain continuous lines. In this section we shall discuss some particular examples of this.

1. *A Line with Coefficients Varying Exponentially.*—In a number of practical problems (exponential electric line, exponential horn for a loud speaker, etc.) one has to deal with continuous lines whose coefficients vary exponentially with the distance. There is a close connection between such exponential lines and our general type of (ϵ_{ij}) line. Let us start from the equations of Sec. 61 and make the following transformation:

$$\begin{aligned} X_1 &= x_1 e^{-\epsilon_{11}z}, & x_1 &= X_1 e^{\epsilon_{11}z}, & x_2 &= X_2 e^{-\epsilon_{11}z}, \\ & & & & X_2 &= x_2 e^{\epsilon_{11}z} \end{aligned} \quad (62.1)$$

$$A_{12} = \epsilon_{12} e^{-2\epsilon_{11}z}, \quad A_{21} = \epsilon_{21} e^{2\epsilon_{11}z}$$

Then our general equations (61.5) yield

$$\left. \begin{aligned} \frac{dX_1}{dz} &= e^{-\epsilon_{11}z} \left(\frac{dx_1}{dz} - \epsilon_{11}x_1 \right) = \epsilon_{12}x_2 e^{-\epsilon_{11}z} = A_{12}X_2 \\ \frac{dX_2}{dz} &= e^{\epsilon_{11}z} \left(\frac{dx_2}{dz} + \epsilon_{11}x_2 \right) = \epsilon_{21}x_1 e^{\epsilon_{11}z} = A_{21}X_1 \end{aligned} \right\} \quad (62.2)$$

These are the equations of propagation of X_1 and X_2 along the exponential line characterized by A_{12} and A_{21} .

The solution obtained in Eq. (61.9) for a wave (x_1, x_2) on the (ϵ_{ij}) line was

$$x_1 = x_{1,0}e^{\pm\gamma z}, \quad x_2 = x_{2,0}e^{\pm\gamma z}$$

where γ is the propagation constant (61.11); hence the solution for the exponential line is

$$X_1 = X_{1,0}e^{(-\epsilon_{11} \pm \gamma)z}, \quad X_2 = X_{2,0}e^{(\epsilon_{11} \pm \gamma)z} \quad (62.3)$$

The plus and minus signs before γ correspond to the two directions of propagation. Thus a finite length of exponential line behaves as a finite length of (ϵ) line terminated by a transformer of ratio $e^{-2\epsilon_{11}z}$.

As an instance of an exponential line, let us consider an electric line with the following values of L and C per unit length:

$$L = L_0e^{-2\epsilon_{11}z}, \quad C = C_0e^{2\epsilon_{11}z}$$

where ϵ_{11} is any real number. Calling X_1 the voltage amplitude and X_2 the current amplitude and dropping the common $e^{i\omega t}$, we obtain the usual line equations

$$\frac{dX_1}{dz} = i\omega L X_2, \quad \frac{dX_2}{dz} = i\omega C X_1$$

which are similar to Eq. (61.7) when $R = G = 0$. These equations of the exponential line have exactly the same form as Eq. (62.2) and possess solutions of the type (62.3). Substituting Eq. (62.3) in the line equations gives

$$\begin{aligned} (-\epsilon_{11} \pm \gamma)X_1 &= i\omega L X_2 \\ i\omega C X_1 &= (\epsilon_{11} \pm \gamma)X_2 \end{aligned}$$

which are two simultaneous linear equations whose determinant must be zero in order to yield a nontrivial solution; hence

$$\gamma^2 = \epsilon_{11}^2 - \omega^2 LC = \epsilon_{11}^2 - \omega^2 L_0 C_0 \quad (62.4)$$

which is exactly the same as Eq. (61.11). The interesting point here is the frequency dependence. The *exponential line* behaves like a *high-pass filter* with a lower cutoff frequency at ω_0 .

$$L_0 C_0 \omega_0^2 = \epsilon_{11}^2$$

since

$$\omega < \omega_0, \quad \gamma^2 = L_0 C_0 (\omega_0^2 - \omega^2), \quad \gamma \text{ real: attenuation}$$

$$\omega > \omega_0, \quad \gamma^2 = -L_0 C_0 (\omega^2 - \omega_0^2), \quad \gamma \text{ pure imaginary: propagation}$$

This result is familiar to electrical engineers and was given here as an example of the general theory. The characteristic impedance of the line varies exponentially as $e^{-2\epsilon_{11}z}$. The phase velocity above cutoff is $1/\sqrt{L_0 C_0 (1 - \omega_0^2/\omega^2)}$.

2. $\epsilon_{11}, \epsilon_{12}, \epsilon_{21}$ *Arbitrary Functions of z.*—As in the first example, we express x_1 and x_2 in terms of ϵ_{11} :

$$x_1 = X_1 e^\varphi, \quad x_2 = X_2 e^{-\varphi} \quad (62.5)$$

where

$$\varphi = \int_0^z \epsilon_{11} dz \quad (62.6)$$

The coefficients A_{12} and A_{21} become

$$A_{12} = \epsilon_{12} e^{-2\varphi}, \quad A_{21} = \epsilon_{21} e^{2\varphi} \quad (62.7)$$

and Eq. (61.5) yields

$$\frac{dX_1}{dz} = A_{12}(z)X_2, \quad \frac{dX_2}{dz} = A_{21}(z)X_1 \quad (62.8)$$

To obtain a four-terminal corresponding to a finite length of the line, we write

$$x_i(z) = b_{i1}(z)x_1(0) + b_{i2}(z)x_2(0) \quad i = 1, 2 \quad (62.9)$$

Increasing the length of dz yields (from the rule for the multiplication of matrices to obtain the effective four-terminal resulting from two four-terminals connected in cascade)

$$b(z + dz) = b(z)b(dz) = [1 + (\epsilon)dz]b(z) \quad (62.10)$$

on substitution of Eq. (61.2). Equation (62.10) may be written explicitly

$$\left. \begin{aligned} \frac{db_{11}}{dz} &= \epsilon_{11}b_{11} + \epsilon_{12}b_{21}, & \frac{db_{12}}{dz} &= \epsilon_{11}b_{12} + \epsilon_{12}b_{22} \\ \frac{db_{21}}{dz} &= \epsilon_{21}b_{11} + \epsilon_{22}b_{21}, & \frac{db_{22}}{dz} &= \epsilon_{21}b_{12} + \epsilon_{22}b_{22} \end{aligned} \right\} \quad (62.11)$$

Equations (62.11) consist of two sets of equations of the type (61.5). The four-terminal equivalent to a length z of the line is obtained by integrating these equations.

3. *A Line with Constant ϵ Coefficients.*—We have already seen that such a line gives propagation with the propagation constant γ of Eq. (61.11) and the characteristic impedances k_1 and k_2 of Eq. (61.13). Then the matrix (b_{ij}) at $z = 0$ is just the unit matrix

$$b(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The matrix elements

$$\left. \begin{aligned} b_{11} &= \frac{1}{k_1 + k_2} (k_2 e^{\gamma z} + k_1 e^{-\gamma z}), \\ b_{12} &= \frac{-k_1 k_2}{k_1 + k_2} (e^{\gamma z} - e^{-\gamma z}) \\ b_{21} &= \frac{-1}{k_1 + k_2} (e^{\gamma z} - e^{-\gamma z}), \\ b_{22} &= \frac{1}{k_1 + k_2} (k_1 e^{\gamma z} + k_2 e^{-\gamma z}) \end{aligned} \right\} \quad (62.12)$$

satisfy the initial conditions and the relation (61.13). Further, (b_{ij}) has determinant one and, in general, represents a finite nonreversible four-terminal. The matrix (ϵ_{ij}) may be found by taking the z derivatives of the matrix (b_{ij}) at $z = 0$.

$$\left. \begin{aligned} \epsilon_{11} &= \frac{k_2 - k_1}{k_2 + k_1} \gamma, & \epsilon_{21} &= \frac{-k_1 k_2}{k_1 + k_2} 2\gamma \\ \epsilon_{12} &= \frac{-1}{k_1 + k_2} 2\gamma, & \epsilon_{22} &= \frac{k_1 - k_2}{k_1 + k_2} \gamma = -\epsilon_{11} \end{aligned} \right\} \quad (62.13)$$

If $k_1 = k_2$, we have $\epsilon_{11} = 0$, which means that the four-terminal is reversible. The line will have zero resistance if

$$\epsilon_{11} \text{ is real; } \quad \epsilon_{12}, \epsilon_{21} \text{ pure imaginary} \quad (62.14)$$

Equation (62.12) is identical with Eq. (55.14) of Chap. IX.

63. Application of Hill's Equation to a Continuous Line

The most general example of a continuous periodic line is given by assuming ϵ_{11} , ϵ_{12} , and ϵ_{21} periodic functions of z , with a common period L . The general wave solution is given by a superposition of the two particular solutions (Floquet's theorem):

$$\left. \begin{aligned} x_1(z) &= e^{\pm \gamma z} f_1(z) \\ x_2(z) &= e^{\pm \gamma z} f_2(z) \end{aligned} \right\} \quad (63.1)$$

where f_1 and f_2 are two periodic function, each with period L . Equations (63.1) are of the same form as the solutions for Mathieu's and Hill's equations, discussed in Chap. VIII. Hill's equation is the appropriate one if we take

$$\begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} = \begin{pmatrix} 0 & c \\ \frac{1}{c} F(z) & 0 \end{pmatrix} \quad (63.2)$$

where $F(z)$ is periodic in z with period L . Mathieu's equation is obtained by taking $F(z)$ a sine or cosine function. Equation (61.5) then becomes

$$\frac{dx_1}{dz} = cx_2, \quad \frac{dx_2}{dz} = \frac{1}{c} F(z)x_1$$

so that

$$\frac{d^2x_1}{dz^2} = F(z)x_1 \quad (\text{Hill}) \quad (63.3)$$

Equation (63.2) implies, evidently, that we are dealing with reversible circuits that may, if desired, contain positive and negative resistance, since $\epsilon_{11} = \epsilon_{22} = 0$ but ϵ_{12} is not pure imaginary. Resistances are avoided by taking $c = i$.

We may reduce more general examples to Hill's equation. To do this, we set, as in Eq. (62.5) to Eq. (62.8)

$$\left. \begin{aligned} \varphi &= \int \epsilon_{11} dz, & \frac{dX_1}{dz} &= A_{12}X_2 \\ A_{12} &= \epsilon_{12}e^{-2\varphi}, & \frac{dX_2}{dz} &= A_{21}X_1 \\ A_{21} &= \epsilon_{21}e^{2\varphi}, \end{aligned} \right\} \quad (63.4)$$

where ϵ_{11} , ϵ_{12} , and ϵ_{21} are assumed periodic with period L in z . Increasing the length of the line by L gives

$$\varphi(z+L) = \varphi(z) + I \quad \text{where } I = \int_0^L \epsilon_{11} dz \quad (63.5)$$

where I will be zero only if ϵ_{11} has an average value of zero. A_{12} and A_{21} will be multiplied by e^{-2I} and e^{2I} , respectively. Then Eqs. (63.4) become

$$\frac{dX_2}{dz} = \frac{d}{dz} \left(\frac{1}{A_{12}} \frac{dX_1}{dz} \right) = A_{21}X_1$$

and if we set

$$z' = \int A_{12} dz \quad (63.6)$$

we obtain

$$\frac{d^2 X_1}{dz'^2} = \frac{A_{21}}{A_{12}} X_1 \quad (63.7)$$

Equation (63.7) is Hill's equation if

$$F = \frac{A_{21}}{A_{12}} = \frac{\epsilon_{21}}{\epsilon_{12}} e^{4\varphi} \quad (63.8)$$

is periodic. This is the case only if I is zero; *i.e.*, only if ϵ_{11} has an average of zero. This means that we may reduce the equations for a periodic line of nonreversible infinitesimal four-terminals to Hill's form only if a section L as a whole appears as a reversible four-terminal. If this is the case, the corresponding continuous periodic line may be reduced to a line of identical reversible four-terminals.

64. Normalization of the Matrix (ϵ_{ij}) and the Pauli Matrices

We have been led in the study of continuous lines to introduce matrices of the form

$$(\epsilon_{ij}) = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & -\epsilon_{11} \end{pmatrix} \quad (64.1)$$

The square of the matrix (ϵ_{ij}) is diagonal.

$$(\epsilon_{ij})^2 = \begin{pmatrix} \epsilon_{11}^2 + \epsilon_{12}\epsilon_{21} & 0 \\ 0 & \epsilon_{11}^2 + \epsilon_{12}\epsilon_{21} \end{pmatrix} = \begin{pmatrix} -|\epsilon_{ij}| & 0 \\ 0 & -|\epsilon_{ij}| \end{pmatrix} \quad (64.2)$$

where $|\epsilon_{ij}|$ is the determinant of (ϵ_{ij}) . We may introduce a normalization factor $E(z)$ so that the square of (ϵ_{ij}) is 1. Assuming this to be done, we may write

$$(\epsilon_{ij})^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad |\epsilon_{ij}| = -1, \\ (\epsilon_{ij}) = \begin{pmatrix} \sqrt{1 - \epsilon_{12}\epsilon_{21}} & \epsilon_{12} \\ \epsilon_{21} & -\sqrt{1 - \epsilon_{12}\epsilon_{21}} \end{pmatrix} \quad (64.3)$$

If we assume (ϵ_{ij}) to be normalized, the factor $E(z)$ will appear explicitly in the other equations.

$$\left. \begin{aligned} \frac{dx_1}{dz} &= E(z)(\epsilon_{11}x_1 + \epsilon_{12}x_2) \\ \frac{dx_2}{dz} &= E(z)(\epsilon_{21}x_1 - \epsilon_{11}x_2) \end{aligned} \right\} \quad (64.4)$$

If we let x represent a column matrix, with elements x_1 and x_2 , we may write Eq. (64.4) symbolically as follows:

$$\frac{dx}{dz} = E(z)(\epsilon_{ij})x \quad (64.5)$$

From Eq. (64.3) it follows that

$$(\epsilon_{ij})^2 = 1; \quad (\epsilon_{ij}) = (\epsilon_{ij})^{-1} \quad (64.6)$$

Hence Eq. (64.5) becomes

$$\frac{1}{E(z)} (\epsilon_{ij}) \frac{dx}{dz} = x \quad (64.7)$$

or

$$\left. \begin{aligned} \frac{1}{E(z)} \left(\epsilon_{11} \frac{dx_1}{dz} + \epsilon_{12} \frac{dx_2}{dz} \right) &= x_1 \\ \frac{1}{E(z)} \left(\epsilon_{21} \frac{dx_1}{dz} - \epsilon_{11} \frac{dx_2}{dz} \right) &= x_2 \end{aligned} \right\} \quad (64.8)$$

From Eq. (64.7) it follows that we may write the operator equation

$$\frac{1}{E} (\epsilon_{ij}) \frac{d}{dz} = 1 \quad (64.9)$$

If we now assume E and the ϵ_{ij} to be constant and not to depend upon z , the matrix operator (ϵ_{ij}) and the differential operator d/dz become independent and hence commute on squaring Eq. (64.9).

$$\left[\frac{1}{E} (\epsilon_{ij}) \frac{d}{dz} \right] \left[\frac{1}{E} (\epsilon_{ij}) \frac{d}{dz} \right] = \frac{1}{E^2} (\epsilon_{ij})^2 \frac{d^2}{dz^2} = \frac{1}{E^2} \frac{d^2}{dz^2} \quad (64.10)$$

Then

$$\frac{1}{E^2} \frac{d^2x}{dz^2} = x \quad (64.11)$$

which is the usual wave equation after the time-dependent part of the solution has been separated out. We assume x_1 and x_2 both periodic in t with the same frequency ω .

We may now generalize to the *three-dimensional wave equation*. We take the three space variables to be z_1 , z_2 , and z_3 and introduce the three constant matrices $(\epsilon)_1$, $(\epsilon)_2$, and $(\epsilon)_3$ and a constant E such that

$$\sum_i (\epsilon)_i \frac{\partial x}{\partial z_i} = Ex \quad (64.12)$$

The operator equation analogous to Eq. (64.9) is

$$\sum_i (\epsilon)_i \frac{\partial}{\partial z_i} = E$$

and squaring yields

$$\sum_i \frac{\partial^2}{\partial z_i^2} = E^2 \quad (64.13)$$

provided that

$$(\epsilon)_i^2 = \delta; \quad (\epsilon)_i(\epsilon)_k + (\epsilon)_k(\epsilon)_i = 0, \quad i \neq k \quad (64.14)$$

The first set of conditions is analogous to the normalization condition (64.3) for one dimension. The second set requires that the matrices anticommute. Three solutions satisfy Eq. (64.14) and are the Pauli matrices

$$(\epsilon)_1 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad (\epsilon)_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (\epsilon)_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad (64.15)$$

It can be shown that any other set of solutions is reducible to a linear combination of the solutions (64.15).

The Pauli matrices are Hermitian symmetrical, a type that we encounter for the first time in this discussion. The *first two* matrices are *reversible*, and the *third is not*. This corresponds to the fact in the Pauli theory of electron spin that these matrices refer to the magnetic moment of an electron having its spin directed along the z_3 axis, a special feature that introduces an asymmetry along the z_3 axis.

The method that we have employed is based on Dirac's method for the linearization of the relativistic wave equation of the electron. Dirac's problem involves four-rowed matrices instead of the two-rowed matrices that we have so far encountered. In the next section we discuss a problem closely connected with Dirac's.

65. Three-phase and Polyphase Lines

Matrices with more than two rows occur in polyphase lines. For instance, a six-terminal inserted in a three-phase line, as in Fig. 65.1, will have four variables at the entrance and an equal number at the exit. We let

$$\left. \begin{array}{ll} x_1 = v & x'_1 = v' \\ x_2 = i & x'_2 = i' \\ x_3 = V & x'_3 = V' \\ x_4 = I & x'_4 = I' \end{array} \right\} \quad (65.1)$$

where the variables are as in Fig. 65.1. Then the matrix for the six-terminal will have four rows and four columns. The elements will be defined by

$$x'_i = \sum_k b_{ik} x_k \quad (65.2)$$

It is a general rule that a $4n$ -terminal in $(n+1)$ parallel lines will require a matrix with $2n$ rows and $2n$ columns.

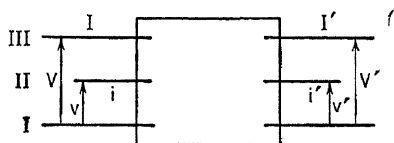


FIG. 65.1.

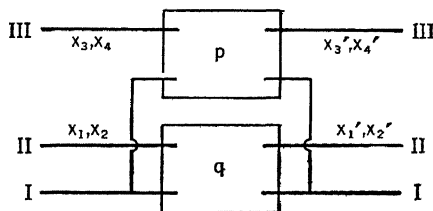


FIG. 65.2.

The determinant of the matrix (b_{ij}) for the six-terminal is no longer unity as for the four-terminal. There are six conditions resulting from the circuit theory, and their form is quite different and leads to consequences different from the four-terminal.

The diagonal form of the matrix (b_{ij}) corresponds to the superposition of two simple waves traveling in opposite directions for the four-terminal. The same general statement is true for the six-terminal except that there will be four simple waves to be superposed instead of two. It does not appear possible, in general, to obtain a simple relation among these four waves.

We shall consider a particular type of six-terminal, built up out of two four-terminals as in Fig. 65.2. Then we may

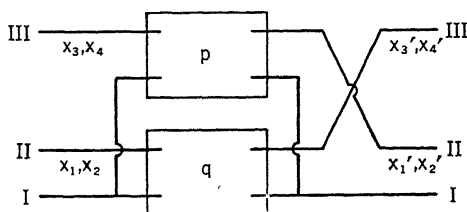


FIG. 65.3.

write Eq. (65.2) in terms of the matrices of the component four-terminals.

$$\left. \begin{aligned} x'_1 &= q_{11}x_1 + q_{12}x_2 & x'_3 &= p_{11}x_3 + p_{12}x_4 \\ x'_2 &= q_{21}x_1 + q_{22}x_2 & x'_4 &= p_{21}x_3 + p_{22}x_4 \end{aligned} \right\} \quad (65.3)$$

so that the matrix for the six terminal is

$$(b_{ij}) = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 \\ q_{21} & q_{22} & 0 & 0 \\ 0 & 0 & p_{11} & p_{12} \\ 0 & 0 & p_{21} & p_{22} \end{pmatrix} \quad (65.4)$$

Another matrix may be obtained for the six-terminal by making the connections as in Fig. 65.3.

$$(b_{ij}) = \begin{pmatrix} 0 & 0 & p_{11} & p_{12} \\ 0 & 0 & p_{21} & p_{22} \\ q_{11} & q_{12} & 0 & 0 \\ q_{21} & q_{22} & 0 & 0 \end{pmatrix} \quad (65.5)$$

If we let the four-terminals making up the six-terminal become infinitesimal, we have

$$\begin{aligned} \begin{pmatrix} p_{11} & p_{22} \\ p_{21} & p_{22} \end{pmatrix} &= \begin{pmatrix} 1 + \pi_{11} dz & \pi_{12} dz \\ \pi_{21} dz & 1 + \pi_{22} dz \end{pmatrix}, \\ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} &= \begin{pmatrix} 1 + \chi_{11} dz & \chi_{12} dz \\ \chi_{21} dz & 1 + \chi_{22} dz \end{pmatrix} \end{aligned} \quad (65.6)$$

so that Eq. (65.4) becomes

$$(b_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + dz \begin{pmatrix} \chi_{11} & \chi_{12} & 0 & 0 \\ \chi_{21} & \chi_{22} & 0 & 0 \\ 0 & 0 & \pi_{11} & \pi_{12} \\ 0 & 0 & \pi_{21} & \pi_{22} \end{pmatrix} \quad (65.7)$$

The four-terminals (and hence the six-terminal) will be reversible if

$$\chi_{11} = \pi_{11} = \chi_{22} = \pi_{22} = 0 \quad (65.8)$$

We may use the expansion (65.7) in Eq. (65.2) to obtain four linear differential equations of the first order for x_1 , x_2 , x_3 , and x_4 . These four equations are of the same form as the Dirac equations for the relativistic electron provided we make the generalization to three dimensions. A set of four-rowed matrices will occur in the theory. There are four possible independent choices for these matrices, and all these matrices are made up of the Pauli matrices. All the matrices are Hermitian and have the following diagonal form:

$$\begin{pmatrix} \gamma_1 & 0 & 0 & . & . & . & . & . \\ 0 & \gamma_2 & 0 & . & . & . & . & . \\ 0 & 0 & \gamma_3 & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & -\gamma_1 & 0 & 0 & . & . \\ . & . & . & 0 & -\gamma_2 & 0 & . & . \\ . & . & . & 0 & 0 & -\gamma_3 & . & . \\ . & . & . & . & . & . & . & . \end{pmatrix} \quad (65.9)$$

In other words, for each wave with propagation constant γ_i , there is a wave with propagation constant $-\gamma_i$. These two waves will propagate in opposite senses.¹

¹ BRILLOUIN, L., *J. phys.*, 7, 401 (1936).

INDEX

A

- Absorption, 174
- Acoustic waves, 158
 - along a pipe of elliptic cross section, 173
- Acoustical branch, 15, 55, 63, 67, 128, 158, 161
- Acoustics, 194
- Adiabatic elastic coefficient, 2
- Admittance, 203
- Ambiguity, in the value of K_1 and K_2 , 104
 - in wave length, 6, 7
- Amplitude $A(r)$, as a periodic function in the direct lattice, 111
- Analogies, 43
 - between electrical and mechanical systems, 40
- Anisotropic medium, 163
 - oscillator, 21
- Antiresonance, 219
- Asymmetric cells, 82, 84
- Atomic planes, 136, 138
- Atomic rows, 136, 138
- Attenuation, 9, 33, 38, 57, 76, 174, 212, 216
- Average perturbation, 112

B

- Baden-Powell, 4, 6, 11, 17, 31
- Bands, absorption, 174
 - four-terminal, 225
 - pass, 19
 - pass-filter, 11, 13, 43
 - passing, 16, 25, 49, 55, 115, 122, 174, 177, 183, 216, 218-221, 225

- Bands, N particles per cell, meaning N , 130
 - stopping, 11, 16, 24, 55, 56, 60, 115, 122, 174, 177, 183, 218, 225, 226
 - structure, 220-223, 225
 - two-terminal, 220-223
- Basis vectors, 131
- Bernoulli, Daniel, 2
- Bloch, F., 140, 171
- Body-centered cubic lattice, 135, 148, 153, 159
- Boltzmann, L., 164
- Born, M., 14, 16, 44, 49, 155, 159, 162, 171
- Bouckaert, L. P., Smoluchowski, R., and Wigner, E., 171
- Boundary conditions, 30, 35, 69, 127, 159
- Bragg's theory, 123
 - of reflection, 113, 115, 118, 120, 122, 136, 138, 191
- Branch, acoustical, 15, 55, 63, 67, 128, 158, 161
 - optical, 15, 55, 63, 67
- Bravais hexagonal lattice, 149, 154
 - notation for crystal planes, 99
- Brillouin, L., 97, 167, 171, 178, 192

C

- Cable, 3, 194
- Campbell, 13, 193, 194
- Canonical form of the line equations, 214
- Capacity, 188, 228
- Carrier wave, 75
- Cauchy, A. L., 3-10
- Cell (cells), 97
 - asymmetric, 82, 84
 - direct and reciprocal, 132

- Cell (cells), elementary, 97, 131
 - of the lattice, 15, 65
 - Change in periodicity, 59
 - Characteristic impedance, 49, 69,
 - 80, 81, 83, 85, 86, 88, 91, 92, 194,
 - 196, 209-211, 213, 214, 217, 218,
 - 224, 226, 229, 231, 232
 - Characteristic temperatures, 164, 166, 171
 - Chodorov, M. I., and Manning, M. F., 171
 - Clairaut, A. C., 3
 - Classical thermodynamics, 166
 - Close-packed hexagonal lattice, 154
 - Close-packed spheres, 149, 150
 - Cloud of electrons, 121
 - Cohesive forces, 168
 - Complete row or plane system, 130
 - Complex amplitude, 147
 - Complex variables, 70
 - Conductance, 228
 - Continuity across the boundary, 127
 - Continuous medium, 32
 - with periodic variations, 106, 108
 - Continuous periodic structure, 130
 - Continuous string, 19
 - with periodic structure, 67, 68
 - Contravariant unit vectors, 97
 - Coupled harmonic oscillators, 58
 - Covariant unit vectors, 97
 - Critical frequency, 9, 13, 14, 18
 - Cross section, 127
 - Crystals, 121, 122, 193
 - Cubic lattice, 152, 153
 - body-centered, 135, 148, 153, 159
 - face-centered, 134, 135, 148, 150,
 - 153, 159
 - Cutoff, 9, 10, 33
 - frequencies, for longitudinal and transverse waves, 164
 - frequency, 19, 32, 38, 162, 164, 230, 231
 - wave length, 162
 - Cyclic conditions, 159, 161
- D
- Davisson, 123
 - De Broglie, M., 64, 123, 158
 - Debye, 162, 164-166, 167, 171
 - Dense spacing, 31
 - Density, 85
 - of points in the lattice planes, 133
 - Determinant, 202, 204, 206, 212, 226, 228, 230, 232, 237
 - infinite, 141, 142, 179
 - Diagonal elements, 212
 - Differential operator, 235
 - Dirac method, 236, 239
 - Direct and reciprocal cells, having reciprocal volumes, 132
 - Direct and reciprocal lattice, 94, 95, 107, 123, 131
 - Direct or reversed pendulum, 190
 - Discontinuities, 130
 - in the ν vs. (a) curve, 106
 - Discontinuous string, 13, 193
 - Dispersion, 10, 69, 70
 - Distance, of interaction, 33
 - between lattice rows or planes, 117, 133
 - Distortion, 70, 76
 - Double line, 39
 - Doubly and triply periodic functions, 99
 - Dulong P. L., and Petit, A. T., 166
- E
- Elastic force, 81
 - Elastic waves, 158, 191, 193
 - Elasticity modulus, 85
 - Electric-circuit theory, 194
 - Electric lines, 69, 193-196
 - continuous, 227
 - loaded, 40, 84
 - Electrical analogue of the one-dimensional diatomic lattice, 47
 - glossary, 195
 - Electromagnetic energy, 40
 - Electromagnetic waves, 158, 193, 194
 - Electron cloud, 127
 - Electron spin, 194, 196, 206, 236
 - Electronic density, 102, 121, 125, 127, 192
 - Electronic waves, 190, 191, 193

- Electrostatic energy, 40
 Elementary cell, 131
 in the direct or reciprocal lattice, 97
 Elementary lattice, 160
 Energy (energies), average potential and kinetic, 92
 density, 72, 73, 89
 equipartition of, 166
 flow of, 69, 72, 73, 85, 86, 89, 90
 of a solid, 164
 velocity, 69, 70, 72, 74, 81, 91
 Entropy, 168
 of a solid body, 167
 Equidistant coupled oscillators, 21
 Equipartition of energy, 166
 Euler, L., 2
 Ewald, P. P., 136, 138
 Expansion, 168
 Exponential horn, 229
 Exponential line, 229, 230
 Extension modulus, 31
 External pressure, 168
- F
- Face-centered cubic lattice, 134, 135, 148, 150, 153, 159
 Filter, band-pass, 193
 high-pass, 13, 14, 19, 36, 42, 230
 low-pass, 9, 13, 19, 36, 37, 40, 42, 84
 Filters, 12, 193, 194
 Finite, 81
 Finite lattice, 28, 34, 69
 Finite piece of crystal, 159
 First-order correction on the frequency, 115
 First zone, 137, 138, 154
 Floquet, 140, 173, 175, 181, 232
 Flow of energy, 69, 72, 73, 85, 86, 89, 90
 Four-terminals, 195, 200, 203, 205, 208, 213, 214, 225, 227, 237
 in cascade, 211, 231
 infinitesimal, 234
 infinitesimal, nonreversible, 228
 inverse, 204
 Four-terminals, nonreversible, 217, 228, 232
 representation of continuous lines, 229
 reversed, 204, 205, 207, 210
 reversible, 205, 211, 214, 216, 217, 226, 228, 229, 232, 234, 239
 Fourier, J. B. J., 77, 130
 series, 2, 3, 35, 143, 184, 185, 191
 multiple, 101, 109, 121
 French, N. R., 49
 Frequency, cutoff, 128
 distribution for waves, 157
 as a periodic function of the wave vector, 104, 142
 Fundamental interval, 25
- G
- Generalized Lagrange coordinates, 196
 Geometrical representation of matrices, 206, 207
 Germer, 123
 Goodwin, T. H., 122
 Group C_2 , 206
 velocity, 70, 74, 75, 80, 91, 160
 of waves, 70
- H
- Hardy, R., 122
 Harmonic content of oscillations, 189
 oscillator, 21, 164, 169
 Hartree, 192
 Heaviside, O., 13
 Hermitian matrices, 236, 239
 Herring, C., 171
 Hexagonal lattice, 123, 149, 150, 153, 154, 158
 close-packed, 150
 first and second zone of, 154
 High-pass filter, 13, 14, 19, 36, 42, 230
 Hill's equation, 143, 178-180, 232, 234
 with a rectangular curve, 180
 Historical background, eighteenth century, 1
 nineteenth century, 3

Homogeneous isotropic medium with
a small periodic perturbation,
137, 139, 143

Hooke, R., 167

Horizontal tangent, 127

Humbert, J., 173, 192

Hund, F., 171

I

Ideal isotropic solid body, 161

Impedances, 203

characteristic (see Characteristic
impedance)

of electric lines, 84, 209

image, 215, 216

iterative, 209

mechanical, 209

Index of refraction, 9, 12

Inductance mutual, 195, 196

Infinite determinant, 141, 142, 179

Infinite line, 81

Infinite number of branches, 130

Infinite set of lattice planes, 132

Infinitesimal portion of the line, 227

Infinitesimal transformation, 227

Inhomogeneous differential equation,
109

Input impedance, 210

Instability and self-excitation, 189

Interaction between molecules, 23

nearest neighbors only, 31, 72, 80

Interatomic distances, 10

Internal pressure, 168

Ions, 192

Isotropic solid, 158

J

Jones, H., 171

Junction, of an aerial telephonic line

with a city cable, 49

of two lattices, 30, 85, 87

K

Kelvin, W. T., 5, 10, 11, 33, 193

Kronecker, L., 67, 141, 143

Kronig, R. de L., 184, 192

L

Lagrange, J. L., 3, 13, 193

Laplace, P. S., 1, 108, 140

Lattice, with basis, 128, 129, 147,
149, 158

body-centered cubic, 135, 148,
153, 159

crystal, 121, 122, 193

cubic, 152, 153

direct, 72, 80, 94, 95, 107, 123,
128, 131, 159

elementary cell in the direct or
reciprocal, 15, 65, 97, 131, 159

finite, 28, 34, 69

hexagonal, 123, 149, 150, 153,
154, 158

of identical particles, 19, 26

NaCl, 44, 49, 52, 55, 154

oblique, 97, 102, 125

one-dimensional, 2, 3, 4, 17, 44,
65, 69

polyatomic, 128

reciprocal, 101, 116, 125, 132,
134, 159

rows and planes, 98, 102, 116, 117,
133, 135, 151

Line, 81

with constants coefficients, 232

electric, 69, 193, 194, 195, 196,
227, 236

of four-terminals, 211, 216

impedance of an electrical, 209

loaded (*see* Loaded line)

Linear combinations, of plane waves,
64

of two exponentials, 113

Linearization of the relativistic
wave equation, 236

Loaded line, 3, 13, 221

with condensers, 222

with inductances, 222

with four-terminals, 224

string, 12, 19, 68

with two-terminals, 218

Longitudinal vibrations, 20, 158,
162, 163, 164

Lorentz, formula of, 11

Loud-speaker, 229
 Low-pass filter, 9, 13, 19, 37
 electrical, 13, 37, 40, 84
 mechanical, 36, 38, 40, 42

M

Magnetic moment, 236
 Manning, M. F., 171
 Mathieu, P., 140, 142, 172-182, 187,
 192, 233
 Matrices, 96, 193, 194, 196, 201, 206,
 207, 213, 218, 220, 226, 227,
 232, 237
 commutative, 209, 218
 complex, 206
 Hermitian symmetrical, 236, 239
 inverse, 202, 206
 multiplication of, 203, 207, 209, 231
 Pauli, 234, 236, 239
 principal axes of, 207, 208, 210, 213
 proper values of, 208, 212, 229
 reversible, 236
 six-terminal, 238
 unit, 96, 204, 206, 227
 Maximum frequencies, 162
 Maxwell's equations of electro-
 magnetism, 70
 Mead, S. P., 49
 Mechanical filter, 12, 36, 38, 40, 42,
 193
 oscillators, 189
 Medium with periodic structure, 158
 Metals, free electrons in, 190
 Mixtures of tiny crystals in, 161
 Modes of vibration, 2
 Modulation, 75
 Molecular structure of matter, 10
 Monatomic lattice, 72, 80, 128, 159
 Morse, P. M., 171, 192
 Mosaic, 125, 129, 130
 Mott, N. F., 171
 Mutual inductance, 195, 196

N

NaCl crystal lattice, 16, 44, 49, 52,
 55, 154
 Newton, I., 1, 3, 31, 32

Normalization condition, 236
 Number of degrees of freedom, 11, 15

O

Oblique coordinate system and lat-
 tice, 97, 102, 125
 Oil films on water, 10
 One-dimensional diatomic lattice, 44
 lattice of polyatomic molecules,
 65
 lattices, 2, 3, 4, 17, 44
 mechanical lattice, 69
 NaCl lattice, 44
 Operator, 235
 Optical branch, 15, 55, 63, 67
 medium, 193
 type, 128, 158
 Orthogonality condition, 110, 112,
 114
 Oscillations of an elliptic lake, 173
 Oscillator harmonic, 21, 164, 166,
 169
 Output impedance, 209

P

Particles interacting at large dis-
 tances, 30
 Passing bands, 11, 16, 25, 49, 55,
 174, 177, 183
 Pauli matrices, 234, 236, 239
 Peierls, R., 171
 Pendulum, 190
 Penney, W. G., 184, 192
 Periodic functions, 99
 Periodic perturbation, 68, 154, 189
 Periodic structure, continuous string
 with, 67, 68, 195
 continuum with, 106, 107, 108
 Perpendicular bisector, 105, 106,
 116, 121, 137, 156, 158
 Perturbation, 35, 68, 108, 115, 119,
 121, 154, 188, 189
 of the lattice, 35
 small, 64, 186, 191
 Phase velocity, 6, 69, 70, 75, 85, 108,
 160
 Planck, M., 164, 169, 190, 191
 Planes, atomic, 136, 138

- Polyatomic lattice, 128
 - Polyphase lines, 236
 - Potential-energy density, 89
 - Poynting's vector, 93, 194
 - Principle of superposition, 2
 - Propagation, of a signal, 74
 - of waves, 70, 175, 193
 - of waves in a continuous two-dimensional medium with a periodic perturbation, 107
 - Propagation vector, 109, 131
 - Proper frequencies, functions, values, and vibrations of, 2, 3, 11, 212, 229
 - Pupin, M. I., 3, 13, 193, 194
- Q
- Quantum mechanics, 196
 - theory of, 164, 169
- R
- Radiation pressure, 167, 168
 - Rayleigh, 167
 - Reciprocal lattice, 94, 95, 108, 123, 130, 136, 139, 142, 151
 - Reciprocity, 77
 - Rectangular lattice, 125
 - Rectifiers, 196
 - Reducing all vectors \mathbf{a} inside the first zone, 154
 - Reflection, 30, 64, 86-88, 91, 207, 213
 - Bragg's, 113, 115, 118, 120, 122, 136, 138, 191
 - at a junction, 217, 218
 - from lattice row, 117
 - partial, 217, 218
 - prevented, 81, 86, 92
 - total, 218
 - of waves, connection between passing or stopping bands and, 64
 - Refractive dispersion, 11
 - Relativistic quantum mechanics, 206
 - Resistance, 38, 217, 225, 228
 - negative, 196
 - Resonance, 12, 219
 - Restricted interval for the wave number, 7, 18
 - Row of diatomic molecules, 22
 - harmonic oscillators, 58
 - lattice points, 28, 123, 136, 138
- S
- Schelkunoff, S. A., 93
 - Schroedinger, E., 190
 - Seitz F., 171
 - Self-consistent field, 192
 - Self-excited oscillator, 186, 187
 - Self-inductance, 228
 - Semiinfinite row of particles, 34
 - Signal, 78, 80
 - with a finite spectrum, 78
 - with an infinite spectrum, 80
 - Six-terminal, 237, 238
 - matrix for, 238
 - Size of atoms, 10
 - Skinne, H. W. B., 171
 - Slater, J. C., 171
 - Smoluchowski, R., 171
 - Space charge, electronic, 121
 - Specific heat, 45, 158, 167
 - Spectrum, 78, 80
 - Spherical first zone, 162, 164
 - Square-centered lattice, 126, 129
 - Stable oscillations, 187
 - Standing wave, 9, 80
 - Statistical computation of entropy, 169
 - Stopping band, 11, 16, 24, 56, 174, 177, 183
 - Straight line in the plane, 98
 - Strain, 32
 - Stratton, J. A., 97, 192
 - String, discontinuous, 13, 193
 - loaded, 12
 - Strutt, M. J. O., 180, 192
 - Superlattice, 29, 35
 - Surface layer, 35
 - Surface tension, 30
 - Surge impedance, 209, 216
 - Symmetric cells, 82, 84
 - Symmetry considerations, 121
 - of the perturbation function, 126

T

- Taylor, B., 2
 Telegraphic cable, 228
 Temperatures, characteristic, 164, 166, 171
 Tensors, 97, 171
 Tides in an elliptic sea, 173
 Thermal expansion, 167
 Three-dimensional lattices, 131
 Three-dimensional wave equation, 235
 Three-phase line, 236
 Transformer ratio, 86, 88
 Transition, from a continuous medium to a discontinuous lattice, 153
 from a diatomic to a monatomic lattice, 57, 62
 near the discontinuity, 118
 Transmission, 88
 Transmission bands (*see* Bands)
 Transverse vibrations, 20, 158, 162, 164
 Triple Fourier series, 136, 139
 Triply periodic functions, 136
 Two-dimensional homogeneous isotropic continuum, 108
 Two-dimensional lattice, 94
 Two-terminal, 219, 220

U

- Unperturbed wave, 115
 Unstable oscillations, 187, 189

V

- Van der Pol, B., 173, 182, 183, 190, 192
 Vaschy, 13
 Velocity, energy, 69, 70, 72, 74, 81, 91
 group, 70, 74, 80, 91, 160
 phase, 9, 69, 70, 75, 85, 108, 160
 of propagation, 1, 4, 6, 19, 20, 31, 32, 35, 38
 Vibrating strings, 2, 3
 Vibrations with an elliptic boundary, 173

- Vincent, 12, 13, 193
 Viscous force, 81

W

- Watson, 177, 178, 192
 Wave, in crystals, 14
 in a discontinuous lattice, 147
 elastic, 158, 191, 193
 in electric filters, 194
 electron, 191, 193
 equation for a two-dimensional continuum, 108, 109
 longitudinal and transverse acoustical, 20, 158, 162, 163, 164
 mechanics, 190, 193, 194
 modulated, 75
 number, 17
 packet, 70
 in a periodic continuous medium, 172
 propagating, in a three-dimensional periodic medium, 139
 propagation, 5, 70, 175, 193, 194, 196, 212, 214, 224, 228, 230, 232
 through the lattice, 137
 Whittaker, 143, 173, 177, 179, 184, 192
 Wigner, E., 155, 171

X

- X-ray reflection from crystal lattices, 102, 115, 118, 120, 121, 158
 X rays, 64, 123, 191, 193

Z

- Zone, 121, 137, 152, 155
 boundary, 142
 first, 105, 126, 129, 137, 138, 154, 158
 second, 106, 129
 structure of, 146, 190
 Zones, for a hexagonal lattice, 124
 in one dimension, 102
 for a square lattice, 126
 in three dimensions, 136, 148
 in two dimensions, 102, 121

W

2849